

Draft Comparative Effectiveness Review

Number xx

Imaging for the Pretreatment Staging of Small Cell Lung Cancer

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No.

Prepared by:

Pending final report

Investigators:

Pending final report

| |
|---|
| <p>This information is distributed solely for the purposes of predissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. The findings are subject to change based on the literature identified in the interim and peer-review/public comments and should not be referenced as definitive. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.</p> |
|---|

AHRQ Publication No. xx-EHCxxx

<Month Year>

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality-enhancement tools, or reimbursement or coverage policies may not be stated or implied.

This report may periodically be assessed for the currency of conclusions. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at: www.effectivehealthcare.ahrq.gov. Search on the title of the report.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is appreciated.

Individuals using assistive technology may not be able to fully access information in this report. For assistance contact [insert program email address].

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation:

Pending final report

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site www.effectivehealthcare.ahrq.gov to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

David Meyers, M.D.
Acting Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Lionel Banez, M.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

Pending final report.

Key Informants

Pending final report.

Technical Expert Panel

Pending final report.

Peer Reviewers

Pending final report.

Imaging for the Pretreatment Staging of Small Cell Lung Cancer

Structured Abstract

Objectives. For small cell lung cancer (SCLC), several imaging modalities can be used to determine cancer staging, which is important to ensure optimal management. Our aim was to synthesize the literature on whether some imaging modalities are better than others for the pretreatment staging of small cell lung cancer. We searched for evidence on comparative accuracy (sensitivity, specificity) as well as subsequent clinical outcomes (choice of treatment, survival, and quality of life).

Data sources. We searched EMBASE, MEDLINE, PubMed, and the Cochrane Library from 2000 through October 15, 2014, for full-length articles on the use of multidetector computed tomography (MDCT), positron emission tomography/computed tomography (PET/CT), magnetic resonance imaging (MRI), combined PET/MRI, endobronchial ultrasound (EBUS), endoscopic ultrasound with fine-needle aspiration (EUS-FNA), and bone scintigraphy in the pretreatment staging of small cell lung cancer.

Review methods. We included studies of pertinent imaging tests on SCLC patients before treatment that reported one or more of the outcomes of interest (studies did not have to directly compare two or more imaging modalities). We extracted data from the included studies and constructed evidence tables. Comparative outcomes of interest included test concordance, staging accuracy (sensitivity and specificity), choice of treatment, timeliness of treatment, tumor response, harms due to overtreatment or undertreatment, survival, and quality of life. For each pair of tests and each assessed aspect (e.g., determination of metastases), we determined whether the evidence was sufficient to permit a conclusion of a difference, a conclusion of similar accuracy, or neither (i.e., insufficient). We rated the risk of bias of individual studies using an internal validity instrument and graded the overall strength of evidence of conclusions using Evidence-Based Practice Center guidance.

Results. The searches identified 2,066 citations; after screening against the inclusion criteria, we included 7 primary studies that enrolled a total of 408 patients. Six of the seven studies were deemed moderate risk of bias, (principally due to failure to report on patient selection, reader blinding to results of comparator tests, and possible spectrum bias), and one was deemed high risk of bias (due to failure to blind readers to results of comparator tests and presence of spectrum bias). One of the studies reported test concordance, three studies reported the comparative accuracy of two or more testing strategies (one of which had also reported test concordance), and four studies reported the accuracy of a single imaging modality. Staging determinations included limited versus extensive disease, osseous (bone or bone marrow) metastases, lymph node involvement, liver metastases, spleen metastases, adrenal metastases, brain metastases, and any distant metastases. The most frequently reported imaging tests were MDCT, [¹⁸F]-fluorodeoxyglucose (FDG) PET/CT, and bone scintigraphy. No studies were included for any other outcomes or for associations with patient comorbidity, body habitus, or tumor characteristics.

Conclusions. Evidence is sparse on imaging modalities in the pretreatment staging of small cell lung cancer. Nevertheless, we drew three conclusions about comparative accuracy: (1) FDG PET/CT is more sensitive than MDCT for detecting osseous metastases; (2) FDG PET/CT is more sensitive than bone scintigraphy for detecting osseous metastases; (3) Standard staging plus FDG PET/CT is more sensitive than standard staging alone for detecting any distant metastases. We assigned a grade of low to the strength of evidence for these conclusions, mostly due to risk of bias and a small number of studies. Research gaps include the dearth of evidence on several tests of interest (particularly MRI, EBUS, EUS, and PET/MRI), a lack of study designs to compare tests on patient-oriented outcomes such as survival, and a lack of data on whether comparative accuracy or effectiveness are associated with patient factors.

Contents

| | |
|---|-------------|
| Executive Summary | ES-1 |
| Introduction..... | 1 |
| Background | 1 |
| Epidemiology | 1 |
| Diagnosis..... | 1 |
| Staging | 1 |
| Imaging Tests..... | 2 |
| Patient Subgroups of Particular Interest | 6 |
| Recent Guidance from Professional Societies and Need for Future Work..... | 6 |
| Scope and Key Questions | 7 |
| Key Questions | 7 |
| Patients, Interventions, Comparisons, and Outcomes..... | 8 |
| Conceptual Framework | 8 |
| Organization of This Report | 9 |
| Methods..... | 10 |
| Topic Development and Refinement | 10 |
| Literature Search Strategy..... | 10 |
| Study Selection | 10 |
| Data Extraction and Management..... | 13 |
| Risk-of-Bias Evaluation..... | 13 |
| Data Synthesis..... | 13 |
| Grading the Body of Evidence for Each Outcome | 14 |
| Peer Review and Public Commentary | 16 |
| Results | 17 |
| Results of Literature Searches | 17 |
| Key Question 1: Concordance and Comparative Accuracy | 19 |
| Test Concordance..... | 19 |
| Comparative Accuracy..... | 20 |
| Key Question 2: Comparative Effectiveness | 23 |
| Key Question 3: Factors Associated with Comparative Outcomes | 23 |
| Discussion..... | 25 |
| Key Findings and Strength of Evidence | 25 |
| Findings in Relationship to What Is Already Known | 28 |
| Implications for Clinical and Policy Decisionmaking | 29 |
| Applicability | 30 |
| Limitations of the Comparative-Effectiveness Review Process | 31 |
| Limitations of the Evidence Base | 31 |
| Research Gaps..... | 32 |
| References | 34 |
| Abbreviations and Acronyms | 37 |

Tables

| | |
|---|----|
| Table 1. Strength of evidence grades and definitions..... | 14 |
| Table 2. Overview of included accuracy data | 20 |
| Table 3. Strength of evidence grades | 26 |

Figures

| | |
|--|----|
| Figure 1. Conceptual framework | 9 |
| Figure 2. Literature flow diagram..... | 19 |
| Figure 3. ROC plots of accuracy data for osseous metastases..... | 22 |

Appendixes

| | |
|--|--|
| Appendix A. Search Strategy | |
| Appendix B. List of Excluded Full Articles | |
| Appendix C. Evidence Tables | |
| Appendix D. Risk of Bias Assessments | |
| Appendix E. Appendix References | |

Executive Summary

Background

Lung cancer is the leading cause of cancer-related mortality, accounting for about 27% of cancer deaths in the United States in 2014.¹ Small cell lung cancer (SCLC) is an aggressive subset of lung cancer characterized by rapid doubling time, high growth fraction, and early development of metastatic disease. This histologic subset of lung cancer is primarily seen in smokers² and comprises approximately 15% of all lung cancers.³ Despite advances in diagnosis, treatment, and management of lung cancer, the 5-year survival rate for SCLC remains dismal at about 6%.¹

Staging involves determining the extent of disease and informs the choice of treatment. SCLC is often staged using the Veterans Administration Lung Study Group (VALSG) system,⁴ which classifies SCLC as either “limited stage” or “extensive stage” disease with the following definitions:

- Limited stage disease (LD): Cancer is confined to one hemithorax and may be present in the regional lymph nodes or in ipsilateral supraclavicular nodes, all of which can be encompassed in a safe radiotherapy field.
- Extensive stage disease (ED): Cancer that cannot be classified as LD, such as when contralateral hilar or supraclavicular nodes, malignant pericardial or pleural effusions, or distant metastatic disease are present.

The National Cancer Institute reported that from 1975–2008, about 70 percent of SCLC cases presented with extensive stage disease, another 21 percent had regional spread such as mediastinal nodal involvement, and only 5 percent were localized (the other 4 percent were unstaged).⁵ The most common sites of metastases for SCLC are the liver, adrenal glands, bone, bone marrow, and brain.⁶

At diagnosis, the majority of SCLC patients have extensive disease, at which time the chance of 5-year survival is only 1 percent.⁷ Chemotherapy has been shown to extend overall survival and improve quality of life. Patients with LD are treated more aggressively with concurrent chemotherapy and radiation with curative intent. After completion of first-line therapy, prophylactic cranial irradiation without evidence of metastases in the brain has been demonstrated to prolong survival in both LD and ED.

Accurate staging of patients is essential to select the optimal treatment plan that will maximize a patient’s chances of survival. On the one hand, overstaging of SCLC risks denying the patient potentially life-saving treatment, while understaging risks subjecting the patient to the unnecessary risk of complications from more aggressive treatment. Given the rapid progression of SCLC, timely diagnosis and staging is important; performing potentially unnecessary tests during the diagnostic and staging process could delay treatment initiation, compromising treatment efficacy.

Multidetector computed tomography (MDCT) of the chest is often one of the first tests performed to diagnose possible lung cancer. For staging SCLC, additional MDCT images are taken of the abdomen, pelvis, or head to detect distant metastases. MDCT has general strengths of widespread availability, high spatial resolution, and high speed and is particularly useful for evaluating the lungs, airways, bowel, and cortical bone. However, because it is a structural imaging modality, it may not detect early metastatic disease involving sites such as the bone marrow or lymph nodes and is not always able to characterize lesions as benign or malignant

based on their morphologic properties. In addition, some patients cannot receive iodinated contrast material due to renal insufficiency, limiting evaluation for presence of hilar lymphadenopathy, vascular abnormalities, and lesion characterization; for these patients, the sensitivity of CT may be lower.

Positron emission tomography (PET) is an imaging modality that localizes the uptake of a positron-emitting radioisotope in the body. [^{18}F]-fluorodeoxyglucose (FDG) is the most commonly used PET radiotracer. Because FDG-PET identifies anatomic sites that harbor metabolically active malignant areas, FDG-PET helps distinguish malignant tumors from benign nodules or masses. FDG-PET can also uniquely detect metabolically active metastases that have not caused anatomic changes. Because PET images lack anatomic detail, combined PET/CT scanners have been developed so the molecular information from PET can be anatomically localized with CT. As of 2014, PET without a concurrent CT is no longer the state of the art.

Magnetic resonance imaging (MRI) is a structural and functional imaging technique that measures the biophysical properties of tissue. MRI has widespread availability, high spatial resolution, and high soft-tissue contrast resolution; this imaging modality is particularly useful for detection and characterization of lesions within tissues even when subcentimeter in size, as well as for evaluation of the internal architecture of organs/tissues such as the brain, spinal cord, breasts, bone marrow, muscles, tendons, ligaments, cartilage, and other solid organs. Also, functional imaging capabilities such as diffusion-weighted imaging and magnetic resonance spectroscopy may be used to improve diagnostic accuracy. MRI examinations take longer to perform and generally cost more than MDCT, and a few patients with implanted electronic or metallic devices or with claustrophobia cannot undergo MRI. Combined PET and MRI scanners are a recent technical development; they promise the sensitivity of PET combined with the anatomic detail of MRI.

Endobronchial ultrasound (EBUS) is a bronchoscopic technique utilizing ultrasonography to visualize structures within and adjacent to the airway wall, whereas endoscopic ultrasound (EUS) is an endoscopic technique that utilizes ultrasonography to visualize structures within and adjacent to the esophageal wall. These techniques are minimally invasive and can be performed on an outpatient basis. EBUS-guided transbronchial needle aspiration (EBUS-TBNA) is generally performed if suspected lymph nodes are in the anterior or superior mediastinum and appear to be accessible based on prior cross-sectional imaging, whereas EUS-guided fine needle aspiration (EUS-FNA) may initially be utilized for nodes that are paraesophageal or subaortic in location or located in the posterior or inferior mediastinum. EBUS-TBNA can also be used to sample hilar lymph nodes. A typical EBUS procedure for lung cancer staging involves standardized sampling of multiple nodal stations that have >5 mm lymph nodes that are detectable and accessible via the EBUS scope.

Bone scintigraphy is a planar molecular imaging technique with widespread availability, high contrast resolution, and relatively low cost compared to FDG-PET/CT. However, false-negative results can occur since bone scintigraphy only indirectly detects the effects of metastatic lesions upon bone turnover. False-positive results can also occur due to visualization of increased bone turnover caused by non-neoplastic etiologies such as fractures and osteomyelitis.

Regarding patient subgroups, performance of various imaging modalities may be affected by comorbidities such as renal insufficiency, which potentially limits use of contrast for MDCT or MRI. Generally, body habitus may limit the diagnostic quality and accuracy for any imaging modality. Many scanners are unable to safely accommodate patients above a particular weight. Tumor characteristics may be associated with comparative accuracy and/or effectiveness.

A 2013 guideline from the American College of Chest Physicians recommended that patients with either proven or suspected SCLC undergo CT of the chest and abdomen or CT of the chest extending through the liver and adrenal glands, as well as MRI of the brain and bone scintigraphy.⁸ In 2014, the American College of Radiology (ACR) appropriateness criteria review gave the highest rating of “usually appropriate” (with regard to staging SCLC) to the following specific modalities: CT of the chest and abdomen with contrast, MRI of the head with and without contrast, and FDG-PET/CT from skull base to mid-thigh.⁹ Bone scintigraphy was rated as “may be appropriate” and considered unnecessary if PET/CT had been performed.

Scope and Key Questions

The scope of this report is to compare imaging modalities in the context of pretreatment staging for SCLC. The Key Questions (KQs) we addressed were as follows:

Key Question 1: What are the test concordance and comparative accuracy of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) for the pretreatment staging of SCLC?

- a. Test concordance
- b. Sensitivity
- c. Specificity
- d. Positive predictive value
- e. Negative predictive value
- f. Positive likelihood ratio
- g. Negative likelihood ratio

Key Question 2: When used for the pretreatment staging of SCLC, what is the comparative effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) on later outcomes?

- a. Choice of treatment (e.g., surgery, chemotherapy, radiation)
- b. Timeliness of treatment
- c. Tumor response
- d. Harms due to overtreatment or undertreatment
- e. Survival
- f. Quality of life

Key Question 3. To what extent are the following factors associated with the comparative accuracy or effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) when used for the pretreatment staging of SCLC?

- a. Comorbidities
- b. Body habitus
- c. Tumor characteristics

Note that two terms above, “accuracy” and “effectiveness”, are used as overarching labels for different sets of outcomes. The “accuracy” outcomes (which are part of Key Question 1), involve an accurate determination of the patient’s stage, whereas the “effectiveness” outcomes (which are listed for Key Question 2), involve the post-staging outcomes such as clinical management and response to treatment. For comparing the *effectiveness* of two imaging modalities, we required that studies make direct comparisons between two or more modalities, whereas for *accuracy*, we included studies that only used one imaging modality. Our full list of inclusion criteria appear in the section below called “Study Selection”.

Methods

Literature Search

With general guidance from the review team, literature searches were performed by medical librarians within the Evidence-Based Practice Center (EPC) Information Center; searches

followed established systematic review methods. We searched the following databases using controlled vocabulary and text words: EMBASE, MEDLINE, PubMed, and the Cochrane Library. The complete search strategy is available in the full report. Each article was screened by at least two people using the database Distiller SR (Evidence Partners, Ottawa, Canada).

For our gray literature searches, we searched for relevant devices on the FDA Web site (i.e., EUS, EBUS, bronchoscopes, CT scanning systems, bone scintigraphy, bone scan). We also browsed ECRI Institute publications including Healthcare Product Comparison Systems reports, Health Technology Forecast and Hotline reports, and ECRI Institute Sourcebase. On the Internet, we searched clinicaltrials.gov, professional organization websites for relevant disease information including prevalence statistics, standards and guidelines, and manufacturer information for relevant diagnostic devices. Professional organization websites were identified via Google and National Guideline Clearinghouse (NGC) searches for relevant SCLC screening/diagnostic/staging guidelines. These Web sites were browsed for disease-specific information: National Comprehensive Cancer Network (NCCN), Society of Nuclear Medicine, and the American College of Radiology.

Study Selection

Our inclusion criteria are listed below in five categories: Publication criteria, study design criteria, patient criteria, test criteria, and data criteria.

Publication Criteria

- a. Full-length articles: The article must have been published as a full-length, peer-reviewed study.
- b. Redundancy: To avoid double-counting patients, in instances in which several reports of the same or overlapping groups of patients were available, only outcome data based on the larger number of patients were included. However, we included data from publications with lower numbers of patients when either (1) a publication with lower patient enrollment reported an included outcome that was not reported by other publications of that study, or (2) a publication with lower patient enrollment reported longer followup data for an outcome.
- c. Publication date: We included studies published since January 1, 2000. Technical progress in all the imaging modalities under consideration means that older studies are likely to underestimate the diagnostic performance of these modalities.
- d. We initially had excluded studies not published in English, but after identifying a relatively low number of qualifying studies, we removed that requirement.

Study Design Criteria

- a. The study must have provided data on at least one test of interest. Ideally, studies would directly compare two or more tests (or test strategies). The comparison may also be addressed indirectly by comparing one set of studies of one imaging test and another set of studies of another imaging test (e.g., a set of studies reporting the accuracy of MRI at mediastinal node staging of SCLC compared to a separate set of studies reporting the accuracy of CT at mediastinal node staging of SCLC). This is an indirect comparison of modalities, so conclusions based on it are weaker.

- b. For comparisons of variants of a given modality, and studies of patient factors or tumor characteristics for KQ3, the comparison must have been planned in advance.
- c. For comparative accuracy (KQ1), the study must have compared both tests to a common reference standard. The reference standard must not have been defined by either imaging test being assessed. We set no requirements on what the reference standard must be (e.g., biopsy, clinical followup).
- d. For comparative effectiveness (KQ2), some patients must have received one of the imaging tests (or test strategies), and a separate group of patients must have received a different imaging test (or test strategy). This design permits a comparison of how the choice of test (or test strategy) might influence choice of treatment, timeliness of treatment, harms due to over/undertreatment, survival, and/or quality of life.
- e. For the influence of patient factors or tumor characteristics (KQ3), the study must have reported comparative accuracy data stratified by patient/tumor factor or comparative-effectiveness data stratified by patient/tumor factor.

Patient Criteria

- a. The study reported data specifically on patients undergoing staging for SCLC, or if the data were combined with other types of patients, at least 85% of the patients in the reported data were undergoing staging for SCLC.
- b. Adults. At least 85 percent of patients must have been aged 18 years or older or data must have been reported separately for those aged 18 years or older.
- c. Studies of the staging of *primary* SCLC were included. Studies for the staging of *recurrent* SCLC were excluded.
- d. Data on imaging tests performed after any form of treatment (e.g., chemotherapy) were excluded, but pretreatment imaging data were considered.

Test Criteria

- a. Type of test. Only studies of the imaging tests of interest were included (listed in the KQs above). Studies of CT that did not explicitly specify whether CT or MDCT was used were assumed to be MDCT. Given our publication date of 2000 and later, we believe it is safe to assume that CT performed in such studies was MDCT.
- b. PET/CT must have been based on a dedicated PET/CT machine. We excluded studies in which PET and CT were acquired separately and later fused for the following reasons. First, methods for image fusion can vary widely. Since accurate staging relies on precise localization of the margin of tumor or metastasis, inclusion of less-precise means of combining PET and CT images would underestimate the performance of PET/CT. Second, image fusion is no longer the state of the art in PET/CT, and since this report is intended to guide use of imaging technologies in the future, it should reflect the technology presently in use.

Data criteria

- a. The study must have reported data pertaining to one of the outcomes of interest (see the KQs).

- For test concordance (part of KQ1), this means reporting the number of patients for whom the two tests provided the same or different results.
 - For comparative accuracy (part of KQ1), this means reporting enough information to calculate both sensitivity and specificity, along with their corresponding confidence intervals (CIs).
 - For comparative choice of treatment (part of KQ2), this means reporting the percentage of patients who received a specific treatment choice for one test or test strategy compared to another test or test strategy.
 - For comparative timeliness of treatment (part of KQ2), this means reporting the duration of time elapsed before the initiation of treatment for one test or test strategy compared to another test or test strategy.
 - For comparative tumor response (part of KQ2), this means reporting the percentage of patients whose tumor responded to treatment for one test or test strategy compared to another test or test strategy.
 - For comparative harms of over- or undertreatment (part of KQ2), this means reporting the percentage of patients who were over- or undertreated (based on authors' judgment) for one test/test strategy compared to another test or test strategy.
 - For comparative survival (part of KQ2), this means either reporting median survival after each imaging test or test strategy, mortality rates at a given time point, or other patient survival such as a hazard ratio.
 - For quality of life (part of KQ2), this means reporting data on a previously tested quality-of-life instrument (such as the Short Form-36) separately for each imaging test or test strategy.
 - For patient factors or tumor characteristics (KQ3), this means reporting data on whether such factors are associated with either comparative accuracy and/or comparative effectiveness.
- b. Regarding the minimum patient enrollment, we required data on at least 10 patients per imaging test or test strategy.
 - c. For all KQs, the reported data must have included at least 50 percent of the patients who had initially enrolled in the study.

Data Extraction and Management

Abstraction forms were constructed in Microsoft Excel, and the data were abstracted into these forms. Elements abstracted include general study characteristics (e.g., country, setting, study design, enrolled N, funding source), patient characteristics (e.g., enrollment criteria, age, sex, final diagnoses including tumor characteristics), details of the imaging methodology (e.g., radiotracer, timing of test, readers, elapsed time between imaging tests, what reference standard was used), risk of bias items, and outcome data. Study methods and outcome data were abstracted by experienced research analysts, in duplicate to ensure accuracy, with discrepancies resolved by consensus. Multiple publications of the same study were grouped as a single study.

Duplicates were identified by examination of author names, study centers, patient enrollment dates, and imaging technologies.

Risk-of-Bias Evaluation

For studies directly comparing two or more imaging tests, we applied a set of nine items involving risk of bias (listed in Appendix D). These items were selected from items in the QUADAS-2 instrument,¹⁰ as well as additional items that specifically address bias in the comparison of imaging tests. For studies of only a single imaging test of interest, the critical issue is whether the study's quantitative estimates could be biased, and we used 14 items that are listed in Appendix D (also based largely on the QUADAS-2 instrument.¹⁰ Each study was assessed by two analysts independently, with disagreements resolved by consensus. Once all individual items were resolved, two analysts assigned each study to a risk-of-bias category (low, moderate, or high), with disagreements resolved by consensus.

Data Synthesis

Decisions about whether meta-analysis was appropriate for a particular data set depended on the judged clinical homogeneity of the different study populations, imaging and treatment protocols, and outcomes. When meta-analysis was not appropriate (due to limitations of reported data), the data were synthesized using a descriptive narrative review approach. We avoided specific numerical thresholds for defining clinical importance of differences because the potential clinical impact of a particular difference in test performance varies according to the particular clinical circumstances of each patient case.

Grading the Body of Evidence for Each Outcome

We used the system for grading evidence on diagnostic tests described in the EPC guidance chapter by Singh et al. (2012).¹¹ This system uses up to eight domains as inputs (study limitations, directness, consistency, precision, reporting bias, dose-response association, all plausible confounders would reduce the effect, and strength of association). The output is a grade for the strength of evidence: high, moderate, low, or insufficient. This grade is provided separately for each outcome of each comparison of each KQ. The grade refers to our confidence in the direction of effect when comparing two imaging technologies, not to the magnitude of the difference between technologies.

A grade was determined separately for each modality comparison (e.g., CT vs. PET/CT). For accuracy, we examined sensitivity and specificity separately. We did not separately grade other accuracy-related outcomes (i.e., predictive values, likelihood ratios) as these grades would be redundant with the grades already assigned for accuracy, since our estimates would be based on the same studies and quantitative syntheses.

A grade of Insufficient was given when evidence did not permit a conclusion for the two modalities being compared with respect to the outcome of interest. For example, if the outcome was comparative sensitivity of CT versus PET/CT, the evidence could support a conclusion that either 1) CT is more sensitive, 2) PET/CT is more sensitive, or 3) the tests are similarly sensitive. If none of these three conclusions could be drawn (as judged by three independent analysts), evidence was graded Insufficient for that comparison. In order to conclude that two modalities differ for an outcome, we used a p value less than 0.05 two-tailed (i.e., the standard value for

alpha). In order to conclude that two modalities are approximately equivalent for an outcome, we used the independent judgment of three analysts (with disagreements resolved by discussion).

If the evidence was sufficient to permit a conclusion, the grade was deemed high, moderate, or low. The grade was provided by three independent raters, with discrepancies resolved by consensus. Specifically, each of three analysts considered all strength of evidence domains listed earlier, and decided on a rating for each evidence base, without knowledge of the ratings of the other two analysts. If any of the three ratings differed, a single rating was reached based on consensus discussion.

Peer Review and Public Commentary

Peer reviewers were invited to provide written comments on the draft report based on their clinical, content, or methodologic expertise. Peer-review comments on the preliminary draft of the report will be considered by the EPC in preparation of the final report. The dispositions of the peer-review comments are documented and will be published 3 months after the publication of the evidence report.

Results

Our searches identified 2,066 citations, of which we excluded 1,851. The most common reasons for exclusion were: studies of other conditions (e.g., non-small cell lung cancer), studies of treatments, and studies not addressing staging. We retrieved the remaining 215 articles, of which we excluded 208. The most common reasons were: studies with fewer than 10 SCLC patients and studies of other conditions. We included the remaining seven publications. Our search of ClinicalTrials.gov identified no additional relevant ongoing studies.

All seven studies were included for KQ1, and none were included for KQs 2 or 3. Two studies were conducted in South Korea, and one each in Japan, Taiwan, Spain, Germany, and Denmark. The only study not published in English was one from Spain.¹² The studies enrolled a total of 408 patients with SCLC. Of the seven studies, three¹³⁻¹⁵ reported the comparative accuracy of two or more tests, and four^{12,16-18} reported single-test accuracy. One¹⁵ of the comparative accuracy studies also reported concordance data.

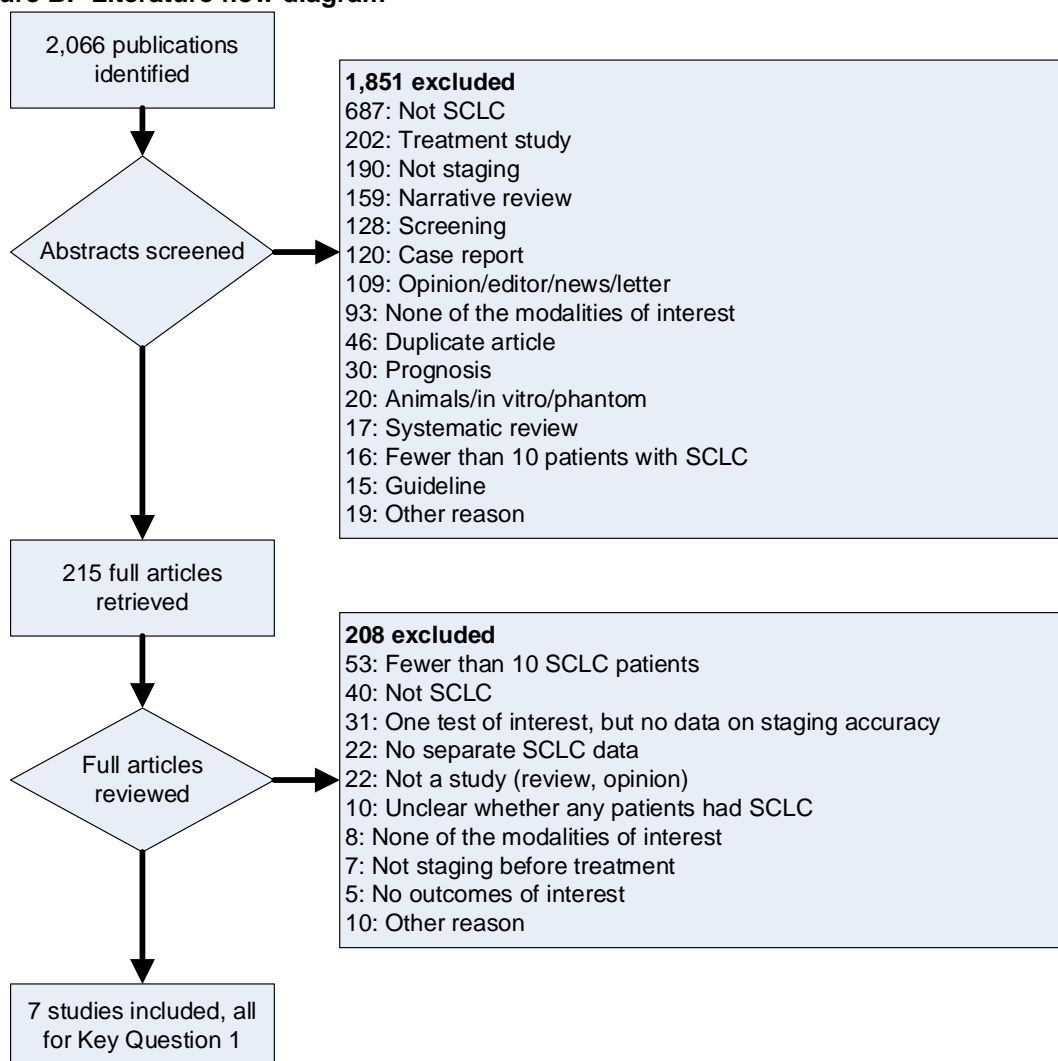
Of the three studies reporting comparative accuracy, we rated three as moderate risk of bias, and one as high risk of bias. The moderate ratings were due to a variety of factors including unknown spectrum bias, failure to report whether test readers had the same clinical information available when interpreting different tests, and the use of test results in determining the reference standard. For instance, Lee et al.¹³ assessed the comparative accuracy of bone scan and FDG-PET for bone marrow metastases. However, the study failed to report whether all patients meeting selection criteria during study period were enrolled. Furthermore, study authors did not specify whether nuclear medicine physicians evaluating FDG-PET images for bone marrow metastases had access to results from the prior bone scan, or additional clinical information which could potentially impact their interpretation. In addition, the study did not specify whether both interpreters had access to the same clinical information before interpreting the images.

The one high risk-of-bias rating was assigned due to the above problems as well as the probability of spectrum bias, failure to report the elapsed time between imaging tests, and clear acknowledgement that test readers had non-complementary knowledge. The four single-test accuracy studies were all rated as moderate risk of bias. Reasons for the moderate rating varied across studies, but common problems included failure to account for inter-reader differences and not blinding the reference standard assessment to test results or other clinical information.

Results of Literature Searches

We depict the literature selection process in Figure A. Searches identified 2,066 citations, of which we excluded 1,851. The most common reasons for exclusion were studies of other conditions (e.g., non-small-cell lung cancer), studies of treatments, and other studies not addressing staging. We retrieved the other 215 articles, of which we excluded 208. The most common reasons were studies with fewer than 10 patients with SCLC and studies of other conditions.

Figure B. Literature flow diagram



Key Question 1: Concordance and Comparative Accuracy

We first briefly summarize test concordance data, then discuss our findings on comparative accuracy.

Test Concordance

One study¹⁵ reported test-test concordance data for three imaging tests (MDCT, FDG PET/CT, and bone scintigraphy). The data appear in Table C-5 of Appendix C. For various

staging determinations (T stage, N stage, pleural effusion, metastases to ipsilateral lung, metastases to contralateral lung, metastases to the liver, metastases to the adrenal glands and metastases to extra-thoracic lymph nodes), authors reported high agreement between MDCT and FDG PET/CT, ranging from 86% to 97%. For the assessment of osseous involvement, however, agreement was lower (83% between MDCT and FDG PET/CT; 46% between MDCT and bone scintigraphy; 57% between FDG PET/CT and bone scintigraphy). The same study also reported the accuracy of these modalities for the assessment of osseous metastases, and we discuss these data (along with all other accuracy data) in the next section.

Comparative Accuracy

An overview of the included accuracy data appears in Table A. Studies reported many different staging determinations (e.g., whether the patient had limited or extensive disease, whether there was metastasis to the brain), but the evidence for any given determination and modality comparison was limited. The largest evidence base involved the comparison of FDG PET/CT to bone scintigraphy for detection of osseous (bone or bone marrow) metastases; this evidence base comprised two studies making direct comparisons (combined n=123) and a single study reporting only bone scintigraphy accuracy data (n=76).

Below, we discuss the results separately for each of eight staging determinations (LD/ED, metastases to osseous structures (bone or bone marrow), lymph node involvement, metastases to adrenal glands, metastases to the liver, metastases to the spleen, any distant metastases, and metastases to the brain). Each of these findings is made even more uncertain by the absence of a consistent and reliable reference method for diagnosis in the studies.

Table A. Overview of included accuracy data

| Staging Determination | Studies Making Direct Comparisons in Accuracy Between Imaging Modalities | Studies Reporting Accuracy Data on a Single Imaging Modality (i.e., for indirect comparisons) |
|--|---|---|
| Limited or extensive disease | Standard staging vs. FDG PET/CT; 1 study; n=28 | Standard staging only; 1 study, n=25 |
| Presence of metastases to osseous structures (bone or bone marrow) | MDCT vs. Bone scintigraphy; 1 study; n=28 MDCT vs. FDG PET/CT; 1 study, n=29 Bone scintigraphy vs. FDG PET/CT; 2 studies, n=123 | Bone scintigraphy only; 1 study, n=76 |
| Presence of lymph node involvement | None | MDCT only; 1 study, n=118 EBUS only; 1 study, n=36 |
| Presence of metastases to adrenal glands | None | MDCT only; 1 study, n=120 |
| Presence of metastases to the liver | None | MDCT only; 1 study, n=120 |
| Presence of metastases to the spleen | None | MDCT only; 1 study, n=120 |
| Presence of any distant metastases | Standard staging vs. standard staging plus FDG PET/CT; 1 study; n=73 | None |
| Presence of metastases to the brain | None | FDG PET/CT only; 1 study, n=21 |

FDG PET/CT=[¹⁸F]-fluorodeoxyglucose positron emission tomography/computed tomography; MDCT=multidetector computed tomography

Limited or Extensive Disease

Two moderate risk-of-bias studies^{15,17} reported data on the ability of imaging tests to determine whether patients with SCLC had LD or ED. Both reported the use of “standard

staging,” which is a combination of multiple testing procedures such as chest x-ray, bone marrow biopsy, and possibly MRI or CT of the brain. In addition, Fischer et al. (2006)¹⁵ reported data on the performance of FDG PET/CT in determining LD/ED. The data suggest that both standard staging and FDG PET/CT had good results (e.g., 86% or 95% for the sensitivity at detecting ED, and specificity of 90% or more at ruling out ED). However, both studies were small (n=28 and n=25, respectively), and the overall data were too imprecise to permit any conclusions about relative accuracy.

Metastases to Osseous Structures (Bone or Bone Marrow)

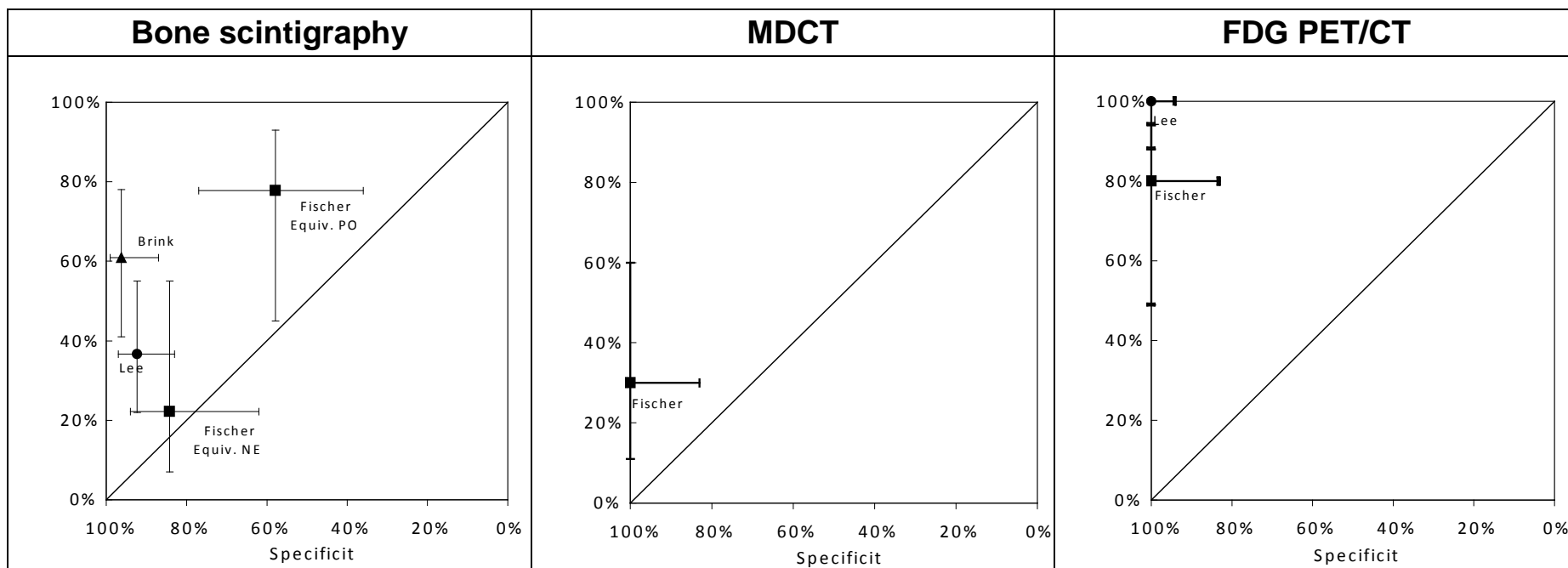
Three moderate risk-of-bias studies reported data on the accuracy of imaging tests to determine whether patients had metastases to osseous structures (bone or bone marrow). One study compared FDG PET/CT to bone scintigraphy;¹³ another study compared CT to FDG PET/CT to bone scintigraphy;¹⁵ the third study provided data only on bone scintigraphy.¹⁸ The accuracy data are shown in Figure B. For bone scintigraphy, the data from Fischer et al. (2006)¹⁵ are plotted twice: once in which equivocal bone scans are treated as positive tests and a second time if equivocal bone scans are treated as negative tests.

Our statistical tests of these data indicated FDG PET/CT was more sensitive than bone scintigraphy in the Lee study, a finding also replicated by the Fischer study if equivocal bone scans were treated as negative tests. If they were considered positive tests, then FDG PET/CT was more specific than bone scintigraphy for the Fischer study. These are direct comparisons. Considered together with the bone scintigraphy results in the Brink study, we concluded that FDG PET/CT is more sensitive than bone scintigraphy for detecting osseous metastases.

Comparing FDG PET/CT to MDCT, only the Fischer study made a direct comparison, and FDG PET/CT was more sensitive but not more specific. Thus, we concluded that FDG PET/CT is more sensitive than MDCT for detecting osseous metastases.

Finally, turning to the comparison of MDCT with bone scintigraphy, treating equivocal bone scans as positive meant a statistical advantage in sensitivity for bone scintigraphy but a statistical advantage in specificity for MDCT. By contrast, if we treated equivocal bone scans as negatives in Fischer, we found no statistical differences in sensitivity or specificity. Other data on bone scintigraphy from Brink and Lee do not suggest marked differences from MDCT. However, the data were too imprecise to permit any conclusions about the comparison of MDCT and bone scintigraphy with respect to osseous metastases.

Figure B. ROC plots of accuracy data for osseous metastases



Note: The three plots show the data on osseous metastases in ROC space. The left plot is for bone scintigraphy, the middle plot is for MDCT, and the right plot is for FDG PET/CT. In each plot, the horizontal axis is specificity (with higher specificity as one moves to the left in the plot), and the vertical axis is sensitivity (with higher sensitivity as one moves up in the plot). The 45 degree line is chance. Thus, an optimal modality would have data in the upper left corner of the plot. Each point is a study (with different studies represented by different shapes). The error bars represent 95% confidence intervals. The study by Brink provided data only on bone scintigraphy, which explains why it does not appear in the other two plots. The study by Lee compared bone scintigraphy and FDG- PET/CT, and the study by Fischer compared all three modalities. The Fischer data are further complicated by the fact that some bone scintigraphy results were considered equivocal by the authors (i.e., neither clearly positive nor clearly negative). Treating equivocal results as positive (as shown with the point labeled “Fischer Equiv. POS”) results in higher sensitivity than treating equivocal bone scan results as negative (as shown with the point labeled “Fischer Equiv. NEG”), but lower specificity. For PET/CT, both studies had point estimates for specificity of 100%, but the confidence intervals around those estimates were too wide to permit conclusions about relative specificity.

Lymph Node Involvement

Two moderate risk-of-bias studies reported data on the accuracy of imaging tests to determine whether patients had lymph node involvement. One study used EBUS for this purpose,¹⁶ and the other used MDCT.¹⁸ The EBUS study reported better accuracy (96% sensitivity and 100% specificity) than the CT study (70% sensitivity, 94% specificity). However, patients in the CT study may have been more difficult to assess for lymph node involvement (as not all lymph nodes were histologically assessed), which would bias the comparison against CT. The indirect nature of the comparison precludes conclusions.

Metastases to Adrenal Glands, Liver, or Spleen

A single moderate-risk-of-bias study¹⁸ reported the single-test accuracy of MDCT for detecting metastases to the adrenal glands, liver, or spleen (separate accuracy data for each of these three types). Because such data were not reported for other imaging modalities by this or other studies, we drew no conclusions about how different modalities compare.

Any Distant Metastases

A single high-risk-of-bias study¹⁴ reported the comparative accuracy of standard staging versus standard staging plus FDG PET/CT for detecting any distant metastases. This study's version of standard staging involved history, physical exam, chest x-ray, chest CT, bone scintigraphy, bone marrow aspiration, and either MRI or CT of the brain. The study reported a large difference in sensitivity (92% for standard staging plus FDG PET/CT vs. only 46% for standard staging), and this difference was statistically significant. The specificities were similar (96% for standard staging plus FDG PET/CT vs. 100% for standard staging), but the precision was too low to permit a conclusion of equivalence on specificity. We deemed the evidence sufficient to permit the conclusion that standard staging plus FDG PET/CT is more sensitive than standard staging alone for detecting any distant metastases. Given that it was only a single high-risk-of-bias study, we rated the strength of the evidence as low.

Metastases to the Brain

A single moderate-risk-of-bias study reported the single-test accuracy of FDG PET/CT of the brain for detecting brain metastases. Because this and others studies did not report such data for other modalities, we drew no conclusions about how different modalities compare in the assessment of brain metastases.

Key Question 2: Comparative Effectiveness

No studies were included for this question.

Key Question 3: Factors Associated with Comparative Outcomes

No studies were included for this question.

Discussion

Key Findings and Strength of Evidence

Based on the evidence we reviewed, we concluded the following:

- FDG PET/CT is more sensitive than bone scintigraphy at detecting osseous metastases (Strength of Evidence: low)

- FDG PET/CT is more sensitive than CT at detecting osseous metastases (Strength of Evidence: low)
- Standard staging plus FDG PET/CT is more sensitive than standard staging alone at detecting any distant metastases (Strength of Evidence: low)

Our strength-of-evidence judgments for these conclusions, along with other evidence we identified, but deemed insufficient to permit conclusions, are listed in Table B below. Note that all three of our conclusions involve the superior sensitivity of FDG PET/CT. We discuss this finding below in the section “Implications for Clinical and Policy Decisionmaking.”

Table B. Strength of evidence grades

| Staging Determination | Test Comparison | Number of Studies and Number of Patients | Study limitations | Direct-ness | Consistency | Precision | Reporting Bias | Magnitude of Effect | SOE Grade* | Conclusion |
|------------------------------|---|---|-------------------|-------------|-------------|-----------|----------------|---------------------|--------------|---|
| LD vs. ED | Standard staging vs. FDG PET/CT | Direct comparison: 1 study ¹⁵ of 28 patients. Indirect comparison: 1 study ¹⁷ of standard staging of 25 patients | Moderate | Mixed | Unknown | Imprecise | None suspected | Not large | Insufficient | NA |
| Osseous metastases | FDG PET/CT vs. bone scintigraphy | Direct comparison: 2 studies ^{13,15} of 123 patients. Indirect comparison: 1 study ¹⁸ of bone scintigraphy of 76 patients | Moderate | Mixed | Consistent | Imprecise | None suspected | Large | Low | FDG PET/CT more sensitive |
| Osseous metastases | CT vs. FDG PET/CT | Direct comparison: 1 study ¹⁵ of 29 patients | Moderate | Direct | Unknown | Imprecise | None suspected | Large | Low | FDG PET/CT more sensitive |
| Osseous metastases | CT vs. bone scintigraphy | Direct comparison: 1 study ¹⁵ of 28 patients. Indirect comparison: 1 study ¹⁸ of bone scintigraphy of 76 patients | Moderate | Mixed | Unknown | Imprecise | None suspected | Not large | Insufficient | NA |
| Lymph node involvement | CT vs. EBUS | Indirect comparison: 1 study ¹⁸ of CT of 118 patients, and 1 study ¹⁶ of EBUS of 36 patients | Moderate | Indirect | Unknown | Imprecise | None suspected | Not large | Insufficient | NA |
| Metastases to adrenal glands | CT vs. anything else | Indirect comparison: 1 study ¹⁸ of CT of 120 patients | Moderate | Indirect | Unknown | Unknown | None suspected | Unknown | Insufficient | NA |
| Metastases to liver | CT vs. anything else | Indirect comparison: 1 study ¹⁸ of CT of 120 patients | Moderate | Indirect | Unknown | Unknown | None suspected | Unknown | Insufficient | NA |
| Metastases to spleen | CT vs. anything else | Indirect comparison: 1 study ¹⁸ of CT of 120 patients | Moderate | Indirect | Unknown | Unknown | None suspected | Unknown | Insufficient | NA |
| Any distant metastasis | Standard staging vs. standard staging plus FDG PET/CT | Direct comparison: 1 study ¹⁴ of 73 patients. | High | Direct | Unknown | Imprecise | None suspected | Large | Low | Standard staging plus FDG PET/CT more sensitive |

Table B. Strength of evidence grades (continued)

| Staging Determination | Test Comparison | Number of Studies and Number of Patients | Study limitations | Direct-ness | Consistency | Precision | Reporting Bias | Magnitude of Effect | SOE Grade* | Conclusion |
|-----------------------|---|---|-------------------|-------------|-------------|-----------|----------------|---------------------|--------------|------------|
| Metastases to brain | FDG PET/CT of the brain vs. anything else | Indirect comparison: 1 study ¹² of FDG PET/CT of 21 patients | Moderate | Indirect | Unknown | Imprecise | None suspected | Unknown | Insufficient | NA |

FDG PET/CT= [18F]-fluorodeoxyglucose positron emission tomography/computed tomography; NA=Not applicable since evidence was insufficient to permit a conclusion for this staging determination for this test-test comparison; SOE=Strength of evidence

* The SOE grade indicates our confidence in the conclusion about the direction of the effect, not about the magnitude of the difference

Implications for Clinical and Policy Decisionmaking

Based on our review of the current evidence, our results suggest two overall conclusions. First, compared to CT and bone scintigraphy (imaging modalities commonly used for staging), FDG PET/CT is more sensitive for detecting osseous metastases in patients with SCLC. Our findings suggest that clinicians evaluating patients for the presence of osseous metastases may consider forgoing bone scintigraphy and routinely using FDG PET/CT instead. Second, for patients with SCLC who have undergone standard staging, the addition of FDG PET/CT increases sensitivity for detecting any distant metastases overall at the patient level.

The evidence base did not allow us to draw conclusions about the comparative specificity of FDG PET/CT compared to these other modalities; thus, we acknowledge that any of the modalities could yield false positives. If a false positive led to inaccurate “upstaging” to extensive disease, a patient might only receive a palliative regimen instead of aggressive chemotherapy aimed at cure. Unfortunately, as we found only 1 study evaluating EBUS and no studies evaluating EUS, MRI, or PET/MRI for inclusion we were unable to assess their comparative accuracy with regard to FDG PET/CT.

Our findings regarding FDG-PET are aligned with recent guidelines from the ACR and ACCP: in 2014, the ACR gave the highest rating of “usually appropriate” to the following specific modalities for staging SCLC: CT of the chest and abdomen with contrast, MRI of the head with and without contrast, and FDG-PET/CT from skull base to mid-thigh.⁹ Bone scintigraphy was rated as “may be appropriate” and considered unnecessary if PET/CT had been performed. Similarly, the 2013 ACCP guideline recommended a staging workup consisting of CT of the chest and abdomen or CT of the chest extending through the liver and adrenal glands, as well as MRI of the brain and bone scintigraphy. For limited disease patients, the guideline “suggested” FDG-PET” as a replacement for bone scan.

SCLC is an aggressive cancer, and timely staging is important to determine treatment decisions based on whether patients have limited or extensive disease. Currently, as part of the standard staging process, patients may undergo bone scintigraphy, CT of the abdomen and pelvis, brain MRI, and FDG PET/CT. Reducing the total number of tests may improve the timeliness of staging and permit faster initiation of treatment.

Higher sensitivity also has other potential important implications for patient care. First, better detection of metastases can improve patient selection for optimal therapy. The higher sensitivity of FDG PET/CT would allow accurate determination of extent of disease, providing clinicians the confidence to offer a comprehensive stage-based treatment plan. Second, earlier detection of extensive disease allows patients to be spared from more aggressive chemotherapy and radiation protocols used for patients with limited disease. Earlier initiation of palliative measures may result in improved quality of life, an important consideration given the current poor prognosis of this disease. Third, improved sensitivity and timeliness of staging may potentially improve the ability of ongoing research trials to prognosticate and detect therapeutic efficacy.

Finally, our results suggest potential resource implications. Although FDG PET/CT may be more expensive than CT and bone scintigraphy, some patients may undergo all three tests when evaluating patients for osseous and other distant metastases outside the brain. Eliminating routine use of bone scintigraphy from SCLC staging protocols in favor FDG-PET/CT could potentially result in some cost savings.

Research Gaps

For characterizing gaps, we used the Hopkins EPC framework proposed by Robinson et al. (2011).¹⁹ This system suggests that reviewers identify a set of important gaps and determine the most important reason for each gap. Each gap should be assigned one of the following reasons for the inability to draw conclusions:

- A. Insufficient or imprecise information: no studies, limited number of studies, sample sizes too small, estimate of effect is imprecise
- B. Information at risk of bias: inappropriate study design; major methodologic limitations in studies
- C. Inconsistency or unknown consistency: consistency unknown (only 1 study); inconsistent results across studies
- D. Not the right information: results not applicable to population of interest; inadequate duration of interventions/comparisons; inadequate duration of followup; optimal/most important outcomes not addressed; results not applicable to setting of interest

We found three primary gaps in the literature on imaging tests for the pretreatment staging of SCLC. The first concerns the dearth of evidence on several tests of interest, particularly MRI, EBUS, EUS, and PET/MRI. This gap exists due to reason A, Insufficient information. Note that we did not restrict our search to studies of particular staging purposes even though some are typically used for specific targets (e.g., brain MRI). EBUS and EUS may be used as much for diagnosis as for staging, which may partially explain the lack of direct evidence on staging SCLC. PET/MRI is a relatively new technology, and we predicted that little would be identified, but future systematic reviews may uncover evidence as it becomes more widespread.

The second gap concerns the absence of study designs to inform the optimal imaging plan for the pretreatment staging of SCLC. Thus, the reason for this gap is reason B, Information at risk of bias. At least three underlying sources exist for the gap: (1) the general lack of direct comparisons of different imaging modalities; (2) the use of mixed reference standards (based on partial histology, other imaging tests, and clinical followup) since not all patients' true stage can be determined before treatment; and (3) the complete lack of studies of comparative effectiveness with patient-oriented outcomes such as management strategy or survival after receiving different test strategies. We recognize that some of these problems are due to the clinical reality that SCLC is fast-growing, necessitating timely initiation of treatment.

The third gap concerns KQ3, which addressed the extent to which comparative accuracy or effectiveness are associated with patient factors (comorbidities, body habitus, tumor characteristics). We identified no studies for this question, so this gap exists due to reason A, Insufficient information. Addressing this problem would be easier than addressing the problems listed in the previous paragraph. Most patient records already contain information on patient characteristics; future research could stratify accuracy results accordingly. Armed with this more detailed data, clinicians and policymakers could possibly recommend tailoring specific strategies for different patient subgroups.

Conclusions

Comparative evidence on the pretreatment staging of SCLC is sparse. We found some low-strength evidence suggesting that FDG PET/CT is more sensitive than CT and bone scintigraphy for assessing osseous metastases.

Executive Summary References

1. American Cancer Society (ACS). Cancer facts & figures 2014. Atlanta (GA): American Cancer Society (ACS); 2014. 70 p.
<http://www.cancer.org/acs/groups/content/%40research/documents/webcontent/acspc-042151.pdf>.
2. American Cancer Society. Cancer prevention & early detection facts & figures 2013. Atlanta (GA): American Cancer Society; 2013. 65 p.
3. National Cancer Institute (NCI). Small cell lung cancer treatment (PDQ) health professional version. Bethesda (MD): National Cancer Institute (NCI), National Institutes of Health (NIH); 2014 Feb 21.
<http://www.cancer.gov/cancertopics/pdq/treatment/small-cell-lung/healthprofessional/page1>. Accessed 2014 Jun 11.
4. Micke P, Faldum A, Metz T, et al. Staging small cell lung cancer: Veterans Administration Lung Study Group versus International Association for the Study of Lung Cancer--what limits limited disease? Lung Cancer. 2002 Sep;37(3):271-6. PMID: 12234695
5. Trends in lung cancer morbidity and mortality. Chicago (IL): American Lung Association; 2012 Apr. 34 p.
6. Glisson BS, Byers LA. Pathobiology and staging of small cell carcinoma of the lung. In: Lilenbaum RC, Nicholson A, Jett JR, editors. UpToDate. Waltham (MA): UpToDate; 2013 May 07.
http://www.uptodate.com/contents/pathobiology-and-staging-of-small-cell-carcinoma-of-the-lung?source=search_result&search=sclc&selectedTitle=2%7E150#H7. Accessed 2014 Jun 12.
7. Shepherd FA, Crowley J, Van Houtte P, et al. The International Association for the Study of Lung Cancer Lung Cancer Staging Project: proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. J Thorac Oncol. 2007 Dec;2(12):1067-77. PMID: 18090577
8. Jett JR, Schild SE, Kesler KA, et al. Treatment of small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013 May;143(5 Suppl):e400S-19S. PMID: 23649448
9. Ravenel JG, Rosenzweig KE, Kirsch J, et al. ACR appropriateness criteria non-invasive clinical staging of bronchogenic carcinoma. J Am Coll Radiol. 2014 Sep;11(9):849-56. PMID: 25086958
10. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003 Nov 10;3(1):25. PMID: 14606960
11. Singh S, Chang S, Matchar DB, et al. Grading a body of evidence on diagnostic tests (AHRQ publication no. 12-EHC079-EF). Chapter 7 of Methods guide for medical test reviews (AHRQ publication no. 12-EHC017). Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2012 Jun 1.
<http://effectivehealthcare.ahrq.gov/ehec/products/246/558/Methods-Guide-for-Medical-Test-Reviews-Full-Guide-20120530.pdf>.
12. Palomar Munoz A, Garcia Vicente AM, Bellon Guardia ME, et al. Is a selective brain (18)F-FDG PET/CT study profitable in patients with small cell lung cancer? Rev Esp Med Nucl Imagen Mol. 2012 May;31(3):124-9. PMID: 21722995
13. Lee JW, Lee SM, Lee HS, et al. Comparison of diagnostic ability between (99m)Tc-MDP bone scan and 18F-FDG PET/CT for bone metastasis in patients with small cell lung cancer. Ann Nucl Med. 2012 Oct;26(8):627-33. PMID: 22729551
14. Sohn BS, Lee DH, Kim EK, et al. The role of integrated 18F-FDG PET-CT as a staging tool for limited-stage small cell lung cancer: a retrospective study. Onkologie. 2012 Aug;35(7):432-8. PMID: 22846975

15. Fischer BM, Mortensen J, Langer SW, et al. A prospective study of PET/CT in initial staging of small-cell lung cancer: comparison with CT, bone scintigraphy and bone marrow analysis. *Ann Oncol*. 2007 Feb;18(2):338-45. PMID: 17060487
16. Wada H, Nakajima T, Yasufuku K, et al. Lymph node staging by endobronchial ultrasound-guided transbronchial needle aspiration in patients with small cell lung cancer. *Ann Thorac Surg*. 2010 Jul;90(1):229-34. PMID: 20609781
17. Shen YY, Shiau YC, Wang JJ, et al. Whole-body 18F-2-deoxyglucose positron emission tomography in primary staging small cell lung cancer. *Anticancer Res*. 2002;22(2):1257-64. PMID: 12168935
18. Brink I, Schumacher T, Mix M, et al. Impact of [(18)F]FDG-PET on the primary staging of small-cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2004 Dec;31(12):1614-20. PMID: 15258700
19. Robinson KA, Saldanha IJ, Mckoy NA. Frameworks for determining research gaps during systematic reviews. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2011 Jun. 79 p. (Methods future research needs report; no.2). http://www.effectivehealthcare.ahrq.gov/ehc/products/201/735/FRN2_Frameworks_20110726.pdf.

Introduction

Background

Epidemiology

Lung cancer is the leading cause of cancer-related mortality, accounting for about 27 percent of cancer deaths in the United States in 2014.¹ In 2011 (the most recent year for which U.S. prevalence data are available), an estimated 402,236 people in the United States were living with lung cancer.² In 2014, approximately 224,210 new cases of lung cancer will be diagnosed.¹ Small cell lung cancer (SCLC) is an aggressive subset of lung cancer that is characterized by rapid doubling time, high growth fraction, and early development of metastatic disease. This histologic subset of lung cancer is primarily seen in smokers³ and composes approximately 15 percent of all lung cancers.⁴ Despite advances in diagnosis, treatment, and management of lung cancer, the 5-year survival rate for SCLC remains dismal at about 6%.¹

The majority of the patients with SCLC (approximately 70%) present with metastatic disease, and about 20 to 30 percent have brain metastases at presentation.^{5,6} Patients with metastatic SCLC have a poor prognosis, with a median survival of only 9–11 months despite aggressive initial treatment. The high rate of distant failure after completion of treatment suggests that systemic micrometastatic spread is common at the time of initial diagnoses but poorly captured by available imaging and surveillance. Platinum-based chemotherapy in combination with etoposide has been the standard of care for over 25 years, with virtually no headway in research. As we pursue novel therapies for this disease, it is imperative that we concurrently pursue strategies that enable us to diagnose and stage these patients' disease in a timely fashion, to guide selection of suitable patients for specific therapy, and to predict clinical outcome.

Diagnosis

SCLC remains a pathologic diagnosis requiring tissue obtained via biopsy. Most cases are diagnosed after patients become symptomatic with cough, shortness of breath, weight loss, or hemoptysis. Imaging, often performed as a first diagnostic step usually reveals a suspicious nodule or mass. Thus, while imaging plays a key role in the diagnostic workup, these tests cannot be used to definitively distinguish SCLC from other types of lung cancer or from other noncancerous conditions (i.e., differential diagnosis). Multidetector computed tomography (MDCT) is widely used as the initial imaging test to determine whether, where, and how to perform a biopsy. Biopsy may be performed via one of several methods (e.g., bronchoscopy, CT-guided percutaneous biopsy, endobronchial ultrasound [EBUS], endoscopic ultrasound [EUS]), depending on the location and size of the mass and patient factors like body mass index and presence of other comorbidities. In a minority of patients, SCLC can also be detected incidentally. In these asymptomatic patients, a chest x-ray or chest CT ordered for other reasons may reveal a nodule or suspicious mass requiring further investigation.

Staging

Staging involves determining the extent of disease and to direct the choice and goals of therapy. SCLC is typically staged using one of two systems, with the Veterans Administration

Lung Study Group (VALSG) system or the less commonly used American Joint Committee on Cancer (AJCC) TNM (T stands for primary tumor, N stands for regional lymph nodes, and M stands for distant metastasis system):

- The VALSG system⁷ classifies SCLC as either “limited stage” or “extensive stage” disease with the following definitions:
 - Limited stage disease (LD): Cancer is confined to one hemithorax and may be present in the regional lymph nodes or in ipsilateral supraclavicular nodes, all of which can be encompassed in a safe radiotherapy field.
 - Extensive stage disease (ED): Cancer that cannot be classified as LD, such as when contralateral hilar or supraclavicular nodes, malignant pericardial or pleural effusions, or distant metastatic disease are present.
- The AJCC TNM system is used less commonly for SCLC but often used for other types of cancer. Lung cancers are classified based on the size of the main tumor, whether it has locally invaded other organs/tissues, spread to lymph nodes, and metastasized to other parts of the body. This information is used to assign a stage between I and IV. A higher stage represents a more extensive spread.

The National Cancer Institute reported that from 1975–2008, about 70 percent of SCLC cases presented with ED, another 21 percent had regional spread such as mediastinal nodal involvement, and only 5% were localized (the other 4% were unstaged).⁸ The most common sites of metastases for SCLC are liver, adrenal glands, bone, bone marrow, and brain.⁹

Some debate exists about which system should be used. The International Association for the Study of Lung Cancer recommended in 2007 that clinicians use the AJCC system.⁵ This recommendation was based on an analysis of more than 8,000 patients with SCLC, which found that both tumor stage and lymph node status were associated with survival. By contrast, others have noted that since approximately two-thirds of patients with SCLC present with advanced disease,⁹ the simpler VALSG system is sufficient to guide treatment. A 2013 guideline from the American College of Chest Physicians recommended that SCLC be staged using both systems.¹⁰ Staging involves evaluating disease in the abdomen, pelvis, bones, and brain.

At diagnosis, the majority of SCLC patients have extensive disease, in whom the chance of survival is low. Chemotherapy has been shown to extend overall survival and improve quality of life. Patients with LD are treated more aggressively with concurrent chemotherapy and radiation with curative intent. After completion of first-line therapy, prophylactic cranial irradiation without evidence of metastases in the brain has been demonstrated to prolong survival in both LD and ED.

Accurate staging of patients is essential to select the optimal treatment plan that will maximize a patient’s chances of survival. On the one hand, overstaging risks denying the patient potentially life-saving treatment, while understaging risks subjecting the patient to unnecessary risk of complications from more aggressive treatment. Given the rapid progression of SCLC, timely diagnosis and staging is important; performing potentially unnecessary tests during the diagnostic and staging process could delay treatment initiation, compromising treatment efficacy.

Imaging Tests

This section discusses several types of imaging tests used to stage SCLC.

Multidetector Computed Tomography

CT scanners acquire cross-sectional images (slices) of the body using x-rays. While early CT scanners could acquire only one slice at a time, current state of the art MDCT scanners can acquire as many as 640 slices at a time. These images can be reconstructed for viewing in any desired plane (multiplanar reconstruction). Since single-detector CT scanners are now obsolete, our report excluded clinical studies using single-detector CT.

Intravenous injection of a radiopaque contrast agent may be performed in some CT examinations. Contrast-enhanced CT can provide additional information about the characteristics of a mass seen in the unenhanced scan, which may facilitate characterization of the mass as probably malignant or probably benign.

MDCT has general strengths of widespread availability, high spatial resolution, and high speed and is particularly useful for evaluating the lungs, airways, bowel, and cortical bone. However, because it is a structural imaging modality, it may not detect early metastatic disease involving sites such as the bone marrow or lymph nodes and is not always able to characterize lesions as benign or malignant based on their morphologic properties. In addition, some patients cannot receive iodinated contrast material due to renal insufficiency, limiting evaluation for presence of hilar lymphadenopathy, vascular abnormalities, and lesion characterization. The major risks of iodinated contrast material include contrast allergic-like reactions (most often mild and self-limited in nature, less commonly moderate or severe requiring therapeutic intervention), contrast-induced nephropathy (most commonly in patients with pre-existing renal insufficiency), and extravasation injury (occurring in up to ~1% of scans, most often in the superficial soft tissues and typically self-limited, rarely leading to compartment syndrome). One potential concern about MDCT is patient exposure to ionizing radiation. However, given the poor prognosis for SCLC when it is typically diagnosed, this concern is relatively unimportant.

MDCT of the chest is often one of the first tests performed to diagnose possible lung cancer. For staging SCLC, additional MDCT images are taken of the abdomen, pelvis, or head to detect distant metastases.

Positron Emission Tomography/Computed Tomography

Positron emission tomography (PET) is an imaging modality that localizes the uptake of a positron-emitting radioisotope in the body. [¹⁸F]-fluorodeoxyglucose (FDG) is the most commonly used PET radiotracer. Because FDG-PET identifies anatomic sites that harbor metabolically active malignant areas, FDG-PET helps distinguish malignant tumors from benign nodules or masses. FDG-PET can also detect metabolically active metastases that may not be detected by structural imaging modalities (e.g., MDCT, magnetic resonance imaging [MRI]).

Because PET images lack anatomic detail, combined PET/CT scanners have been developed so the molecular information from PET can be anatomically localized with CT. As of 2014, PET without a concurrent CT is rarely performed. Therefore, our analysis of PET data included only studies in which patients were imaged using a combined PET/CT scanner.

FDG-PET/CT is sometimes available (but less often than CT) and has high contrast resolution. Because it is quantitative, measurements of the specific uptake in individual lesions or global disease burden may be performed. However, FDG uptake within a lesion does not always indicate malignancy. Infectious and non-infectious inflammatory lesions may also demonstrate FDG uptake, potentially leading to false-positive results. Furthermore, some small or indolent malignant lesions (such as lung adenocarcinoma in situ or carcinoid tumor) may have little to no radiotracer uptake, potentially leading to false-negative results. Lastly, patients with

elevated serum glucose levels (greater than 200 mg/dl) do not generally undergo FDG-PET/CT, given the potential for false-negative results. Again, one potential concern about MDCT is patient exposure to ionizing radiation. However, given the poor prognosis for SCLC when it is typically diagnosed, this concern is relatively unimportant.

Magnetic Resonance Imaging

MRI uses magnetic fields and radio waves to generate three-dimensional images of the body. Unlike PET and CT, MRI does not use ionizing radiation and thus poses no radiation-related risks to the patient. Paramagnetic contrast agents can be administered during the MRI examination to give additional information about the nature of a mass in the same way that iodinated contrast agents can enhance a CT scan.

MRI is a structural and functional imaging technique with widespread availability, high spatial resolution, and high soft-tissue contrast resolution; this imaging modality is particularly useful for detecting and characterizing lesions within tissues even when subcentimeter in size, as well as for evaluating the internal architecture of organs/tissues such as the brain, spinal cord, breasts, bone marrow, muscles, tendons, ligaments, cartilage, and other solid organs. Also, functional imaging capabilities such as diffusion-weighted imaging and magnetic resonance spectroscopy may be used to improve diagnostic accuracy. MRI examinations take longer to perform and generally cost more than MDCT, and some patients with implanted electronic or metallic devices or with claustrophobia cannot undergo MRI.

Positron Emission Tomography/Magnetic Resonance Imaging

Combined PET and MRI scanners are a recent technical development: they promise the sensitivity of PET combined with the anatomic detail of MRI. PET/MRI is a hybrid molecular/structural imaging technique that possesses the general strengths provided by PET and MRI listed above, along with a lower radiation dose compared to PET/CT, and potentially improved PET quantification and motion compensation. However, in addition to the weaknesses of PET and MRI listed above, it is not widely available; not currently reimbursed for by insurance companies; more expensive in terms of instrumentation than PET/CT, MDCT, and MRI; involves longer examination times; and requires additional training of personnel in terms of safety, protocol optimization, and study interpretation.

Endobronchial Ultrasound and Endoscopic Ultrasound

EBUS is a bronchoscopic technique utilizing ultrasonography to visualize structures within and adjacent to the airway wall, whereas EUS is an endoscopic technique that utilizes ultrasonography to visualize structures within and adjacent to the esophageal wall. These techniques are minimally invasive and can be performed on an outpatient basis. Patients with suspected spread of lung cancer to mediastinal lymph nodes may undergo preoperative (or intraoperative) EBUS-guided biopsy and/or EUS-guided biopsy for real-time, minimally invasive pathologic mediastinal N staging, even for lymph nodes that are subcentimeter in size or located near blood vessels. EBUS is used to sample hilar and mediastinal lymph nodes that surround the tracheal bronchial tree, whereas EUS can be used to sample other lymph nodes in proximity to the esophagus. These techniques are utilized during the staging workup of patients with lung cancer (most often non-small cell lung carcinoma [NSCLC]), as accurate N staging is important to determine the optimal therapeutic approach (i.e., surgical vs. nonsurgical therapy),

especially since the sensitivity and specificity of CT, MRI, and FDG-PET/CT for noninvasive N staging of lung cancer are not 100%.

EBUS-guided transbronchial needle aspiration (EBUS-TBNA) is generally performed if suspected lymph nodes are in the anterior or superior mediastinum and appear to be accessible based on prior cross-sectional imaging, whereas EUS-guided fine-needle aspiration (EUS-FNA) may initially be utilized for nodes that are paraesophageal or subaortic in location or located in the posterior or inferior mediastinum. EBUS-TBNA can also be used to sample hilar lymph nodes.

EBUS can also be used to inform T staging of lung cancer and may potentially improve the determination of the distance between the tumor and the carina, which is a determinant of T stage. Furthermore, it may help distinguish between airway invasion by tumor compared to airway compression by tumor. EBUS is also used for diagnostic evaluation of endobronchial lesions, peripherally located pulmonary nodules, and mediastinal lesions, as well as for guiding endobronchial therapy.

Bone Scintigraphy

Bone scintigraphy uses a gamma camera or SPECT (Single Photon Emission Computed Tomography) scanner to create two-dimensional images of the distribution of a radiotracer, typically technetium-99m methylene diphosphonate (Tc99m-MDP). Scintigraphy of radiotracers that localize to the bones is useful in cancer staging because areas of high radiotracer uptake could represent metastases; however, uptake occurs with other common conditions as well, meaning that false-positive findings are not unusual.

Bone scintigraphy is a two-dimensional molecular imaging technique with widespread availability, high contrast resolution, and relatively low cost compared to FDG-PET/CT. However, false-negative results can occur, particularly when lytic osseous lesions are present, as bone scintigraphy indirectly reveals sites of metabolic bone turnover in reaction to bone marrow metastatic disease rather than the metastatic lesions directly. False-positive results can also occur due to other nonmalignant conditions that lead to increased bone turnover, such as fractures and osteomyelitis. Also, this technique has only planar capability (compared to tomographic capability of the other imaging techniques described above), further limiting its diagnostic capability. Lastly, bone scintigraphy is less quantitative than PET/CT and does not allow evaluation of non-osseous structures of the body (in contrast to MDCT, MRI, and PET).

Test Utilization and Costs

In 2008, the combined annual imaging days per person-year were 2.3 CT, 0.49 MRI, 0.70 PET, and 0.13 bone scintigraphy. The annual rates of imaging from 2004 to 2008 increased 0.5% for CT, 3.2% for MRI, and 18.0% for PET, and decreased 12.7% for bone scintigraphy. PET continues to grow rapidly with evidence that it is replacing bone scintigraphy without a decline in CT.¹¹ Regarding PET or PET/CT, an estimated 1.62 million clinical PET and PET/CT scans were performed in 2014, representing a net decrease of about 13% compared to 2011.¹² About 94% of all PET studies in 2012 were for cancer (19% for diagnosis, 38% for staging, 13% for treatment planning, and 30% for follow up).¹³ Regarding combined MRI/CT, utilization rates have increased from 64.3 to 109.1 per 1,000 person years from 2000 to 2009.¹⁴ However, growth has slowed in recent years, with an average annual decline in the imaging growth rate of 4.7% between 2000 and 2009.¹⁴

Regarding costs, we searched the CMS website (<http://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx>) for Medicare rates (national payment amounts, based on 2015 CPT codes and CMS.gov physician fee schedule) for specific procedures and found the following:

- CT chest with contrast: \$229.55
- CT abdomen and pelvis with contrast: \$312.50
- CT head/ brain wo contrast: \$116.20
- FDG-PET/CT (skull base to thigh): \$1523.19
- MRI chest w/wo contrast: \$581.37
- MRI abdomen w/wo contrast: \$508.43
- MRI pelvis w/wo contrast: \$507.36
- MRI head (or brain) w/wo contrast: \$378.28
- Bone scintigraphy: \$261.37
- Bronchoscopy with EBUS/TBNA: \$281.39

Patient Subgroups of Particular Interest

We are unaware of patient factors that might differentially affect the accuracy of imaging modalities specifically for SCLC. As previously noted, performance of various imaging modalities may be affected by comorbidities such renal insufficiency, which potentially limits the use of contrast for MDCT. Generally speaking, body habitus may limit the diagnostic quality and accuracy for any imaging modality. Many scanners are unable to safely accommodate patients above a particular weight. However, these are general patient considerations, not specific to the use of imaging for SCLC.

Recent Guidance from Professional Societies and Need for Future Work

In 2014, the American College of Radiology (ACR) appropriateness criteria review gave the highest rating of “usually appropriate” (with regard to staging SCLC) to the following specific modalities: CT of the chest and abdomen with contrast, MRI of the head with and without contrast, and FDG-PET/CT from skull base to mid-thigh.¹⁵ Bone scintigraphy was rated as “may be appropriate” and considered unnecessary if PET/CT had been performed. ACR noted that PET/CT is often helpful in staging SCLC and may result in a change in staging in up to 17 percent of cases, mostly cases in which PET/CT detects extensive disease that was not detected by other modalities. It may also detect additional involved lymph nodes, leading to revisions in treatment plans for patients scheduled to receive radiotherapy. As PET/CT has already become frequently adopted as part of the SCLC staging process, it is important to establish whether evidence exists to support this practice.

Similarly, a 2013 guideline from the American College of Chest Physicians recommended that patients with either proven or suspected SCLC undergo CT of the chest and abdomen or CT of the chest extending through the liver and adrenal glands, as well as MRI of the brain and bone scintigraphy.¹⁰ Furthermore, in patients thought to have limited disease, the guideline suggested FDG-PET/CT imaging to improve detection of metastases (with the potential to increase the stage and change the treatment choice). However, the evidence underlying this recommendation of PET for potential upstaging is limited and considered weak. The last search date of the guideline was 2011; therefore, three years of additional literature were available for our report

(2012 through October 2014 when searches were performed). A primary objective of this report is to update the evidence base by including more recent studies.

Scope and Key Questions

Key Questions

Key Question 1: What are the test concordance and comparative accuracy of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) for the pretreatment staging of SCLC?

- a. Test concordance
- b. Sensitivity
- c. Specificity
- d. Positive predictive value
- e. Negative predictive value
- f. Positive likelihood ratio
- g. Negative likelihood ratio

Key Question 2: When used for the pretreatment staging of SCLC, what is the comparative effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) on later outcomes?

- a. Choice of treatment (e.g., surgery, chemotherapy, radiation)
- b. Timeliness of treatment
- c. Tumor response
- d. Harms due to overtreatment or undertreatment
- e. Survival
- f. Quality of life

Key Question 3. To what extent are the following factors associated with the comparative accuracy or effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) when used for the pretreatment staging of SCLC?

- a. Comorbidities
- b. Body habitus
- c. Tumor characteristics

Note that two terms above, “accuracy” and “effectiveness”, are used as overarching labels for different sets of outcomes. The “accuracy” outcomes (which are part of Key Question 1), involve an accurate determination of the patient’s stage, whereas the “effectiveness” outcomes (which are listed for Key Question 2), involve the post-staging outcomes such as clinical management and response to treatment. For comparing the *effectiveness* of two imaging modalities, we required that studies make direct comparisons between two or more modalities, whereas for *accuracy*, we included studies that only used one imaging modality. Our full list of inclusion criteria appear in the section below called “Study Selection”.

Patients, Interventions, Comparisons, and Outcomes

Populations

- Adult patients with known SCLC or combined SCLC who are undergoing imaging test(s) for staging and who have not yet received treatment.

Interventions

Imaging using one or more of the following tests:

- Multidetector computed tomography (MDCT)
- PET/CT
- MRI
- PET/MRI
- EBUS
- EUS-FNA
- Bone scintigraphy

Comparators

- Any direct comparisons of the imaging tests of interest
- Any combination of the imaging test of interest

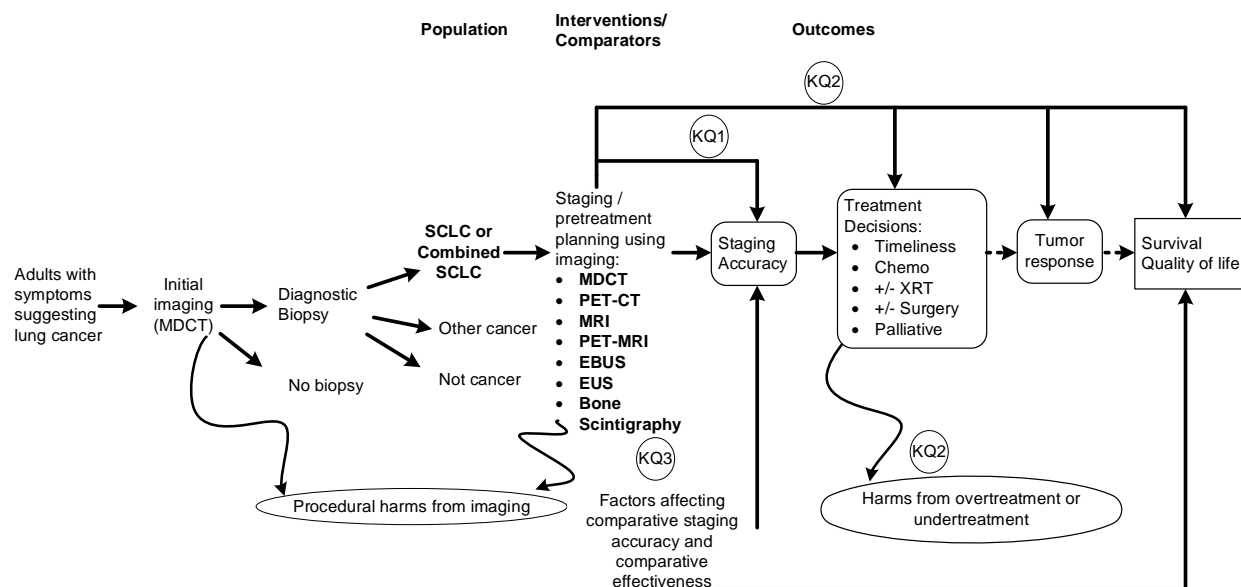
Outcomes

- Test concordance
- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value
- Positive likelihood ratio
- Negative likelihood ratio
- Choice of treatment (e.g., surgery, chemotherapy, radiation)
- Timeliness of treatment
- Tumor response
- Harms due to overtreatment or undertreatment
- Survival
- Quality of life

Conceptual Framework

The conceptual framework for this review appears in Figure 1. This figure depicts the key questions within the context of the PICOTS listed in the previous section. The figure shows that the populations of interest (patients with either SCLC or Combined SCLC) are staged using various imaging modalities, which leads to choices in patient management, leading (hopefully) to tumor response, and finally to patient-oriented outcomes of survival and quality of life. Key harms of interest are the harms due to overtreatment or undertreatment.

Figure 1. Conceptual framework



Organization of This Report

In the remaining three chapters of this report, we present the methods for this systematic review, the results for each key question, and a discussion of the findings. Within the Results chapter, we provide the results of the literature searches and selection procedures, then the results for Key Question 1. The Discussion section provides an overview of our findings and how they relate to what is already known. In that section, we also discuss clinical and policy decisionmaking, the applicability of the evidence, limitations of our review as well as limitations of the evidence we reviewed, and any major gaps in existing research.

Methods

Topic Development and Refinement

Initially a panel of key informants gave input on the Key Questions (KQs) to be examined; these KQs were posted on AHRQ's Web site for public comment between September 4, 2014, and September 24, 2014, and revised as needed. We then drafted a protocol for the Comparative Effectiveness Review and recruited a panel of technical experts to provide high-level content and methodologic expertise throughout the development of the review. The protocol is registered with PROSPERO (<http://www.crd.york.ac.uk/PROSPERO/>) as review # CRD42014015429.

Literature Search Strategy

Literature searches were performed by medical librarians within the Evidence-Based Practice Center (EPC) Information Center; searches followed established systematic review methods. We searched the following databases using controlled vocabulary and text words: EMBASE, MEDLINE, PubMed, and the Cochrane Library. The following grey literature sources were searched using text words: ClinicalTrials.gov, Centers for Medicare & Medicaid Medicare Coverage Database, ECRI Institute Health Devices, Healthcare Standards, Internet, Medscape, National Guideline Clearinghouse™, and the U.S. Food and Drug Administration. The search strategy appears in Appendix A.

Literature screening was performed by experienced research analysts using the database Distiller SR (Evidence Partners, Ottawa, Canada). Literature search results were initially screened for relevancy. Relevant abstracts were screened against the study inclusion criteria in duplicate. Studies that appeared to meet the study inclusion criteria were retrieved in full and screened in duplicate against the study inclusion criteria. All disagreements were resolved by consensus discussion among the two original screeners or a third screener.

The literature searches will be updated during the peer-review process before finalization of the review.

Study Selection

Our inclusion criteria are listed below in five categories: Publication criteria, study design criteria, patient criteria, test criteria, and data criteria.

Publication Criteria

- a. Full-length articles: The article must have been published as a full-length, peer-reviewed study.
- b. Redundancy: To avoid double-counting patients, in instances in which several reports of the same or overlapping groups of patients were available, only outcome data based on the larger number of patients were included. However, we included data from publications with lower numbers of patients when either (a) a publication with lower patient enrollment reported an included outcome that was not reported by other publications of that study, or (b) a publication with lower patient enrollment reported longer followup data for an outcome.

- c. Publication date: We included studies published since January 1, 2000. Technical progress in all the imaging modalities under consideration means that older studies are likely to underestimate the diagnostic performance of these modalities.
- d. We initially had excluded studies not published in English, but after identifying a relatively low number of qualifying studies, we removed that requirement.

Study Design Criteria

- a. The study must have provided data on at least one test of interest. Ideally, studies would directly compare two or more tests (or test strategies). The comparison may also be addressed indirectly by comparing one set of studies of one imaging test and another set of studies of another imaging test (e.g., a set of studies reporting the accuracy of MRI at mediastinal node staging of SCLC compared to a separate set of studies reporting the accuracy of CT at mediastinal node staging of SCLC). This is an indirect comparison of modalities, so conclusions based on it are weaker.
- b. For comparisons of variants of a given modality, and studies of patient factors or tumor characteristics for KQ3, the comparison must have been planned in advance.
- c. For comparative accuracy (KQ1), the study must have compared both tests to a common reference standard. The reference standard must not have been defined by either imaging test being assessed. We set no requirements on what the reference standard must be (e.g., biopsy, clinical followup).
- d. For comparative effectiveness (KQ2), some patients must have received one of the imaging tests (or test strategies), and a separate group of patients must have received a different imaging test (or test strategy). This design permits a comparison of how the choice of test (or test strategy) may influence choice of treatment, timeliness of treatment, harms due to over-/undertreatment, survival, and/or quality of life.
- e. For the influence of patient factors or tumor characteristics (KQ3), the study must have reported comparative accuracy data stratified by patient/tumor factor or comparative-effectiveness data stratified by patient/tumor factor.

Patient Criteria

- a. The study reported data specifically on patients undergoing staging for SCLC, or if the data were combined with other types of patients, at least 85% of the patients in the reported data were undergoing staging for SCLC.
- b. Adults. At least 85 percent of patients must have been aged 18 years or older, or data must have been reported separately for those aged 18 years or older.
- c. Studies of staging for *primary* SCLC were included. Studies of staging for *recurrent* SCLC were excluded.
- d. Data on imaging tests performed after any form of treatment (e.g., chemotherapy) were excluded, but pretreatment imaging data were considered.

Test Criteria

- c. Type of test. Only studies of the imaging tests of interest were included (listed in the key questions above). Studies of CT that did not explicitly state (or we could not determine)

whether they used CT or MDCT were assumed to be MDCT. Given our publication date of 2000 and later, we believe it is safe to assume that CT performed in such studies was MDCT.

- a. PET/CT must have been based on a dedicated PET/CT machine. We excluded studies in which PET and CT were acquired separately and later fused for the following reasons. First, methods for image fusion can vary widely. Since accurate staging relies on precise localization of the margin of tumor or metastasis, inclusion of less precise means of combining PET and CT images would underestimate the performance of PET/CT. Second, image fusion is no longer the state of the art in PET/CT, and since this report is intended to guide use of imaging technologies in the future, it should reflect the technology presently in use.

Data Criteria

- a. The study must have reported data pertaining to one of the outcomes of interest (see the KQs).
 - For test concordance (part of KQ1), this means reporting the number of patients for whom the two tests provided the same or different results.
 - For comparative accuracy (part of KQ1), this means reporting enough information to calculate both sensitivity and specificity, along with their corresponding confidence intervals (CIs).
 - For comparative choice of treatment (part of KQ2), this means reporting the percentage of patients who received a specific treatment choice for one test or test strategy compared to another test or test strategy.
 - For comparative timeliness of treatment (part of KQ2), this means reporting the duration of time elapsed before the initiation of treatment for one test or test strategy compared to another test or test strategy.
 - For comparative tumor response (part of KQ2), this means reporting the percentage of patients whose tumor responded to treatment for one test or test strategy compared to another test or test strategy.
 - For comparative harms of overtreatment or undertreatment (part of KQ2), this means reporting the percentage of patients who were overtreated or undertreated (based on authors' judgment), for one test/test strategy compared to another test/test strategy.
 - For comparative survival (part of KQ2), this means either reporting median survival after each imaging test or test strategy, or mortality rates at a given time point, or other patient survival such as a hazard ratio.
 - For quality of life (part of KQ2), this means reporting data on a previously tested quality-of-life instrument (such as the Short-Form 36) separately for each imaging test or test strategy.
 - For patient factors or tumor characteristics (KQ3), this means reporting data on whether such factors are associated with either comparative accuracy and/or comparative effectiveness.

- b. Regarding the minimum patient enrollment, we required data on at least 10 patients per imaging test or test strategy.
- c. For all KQs, the reported data must have included at least 50 percent of the patients who had initially enrolled in the study.

Data Extraction and Management

Abstraction forms were constructed in Microsoft Excel, and the data were abstracted into these forms. Elements abstracted included general study characteristics (e.g., country, setting, study design, enrolled N, funding source), patient characteristics (e.g., enrollment criteria, age, sex, final diagnoses including tumor characteristics), details of the imaging methodology (e.g., radiotracer, timing of test, readers, elapsed time between imaging tests, what reference standard was used), risk-of-bias items, and outcome data. Outcome data were dual abstracted to ensure accuracy, with discrepancies resolved by consensus. Multiple publications of the same study were grouped as a single study. We identified these by examining author names, study centers, patient enrollment dates, and imaging technologies.

Risk-of-Bias Evaluation

For studies directly comparing two or more imaging tests, we applied a set of nine items involving risk of bias (listed in Appendix D). We devised these items after considering the QUADAS-2 instrument,¹⁶ as well as additional issues that specifically address bias in the comparison of imaging tests. For studies of only a single imaging test of interest, the critical issue was whether the study's quantitative estimates could be biased, and we used 14 items, which are listed in Appendix D (also based on considering the QUADAS-2 instrument.)¹⁶ Two analysts independently assessed each study, with disagreements resolved by consensus. Once all individual items were resolved, two analysts assigned each study to a risk-of-bias category (low, moderate, or high), with disagreements resolved by consensus.

Data Synthesis

Decisions about whether meta-analysis is appropriate depended on the judged clinical homogeneity of the different study populations, imaging and treatment protocols, and outcomes. When meta-analysis was not appropriate (due to limitations of reported data), the data were synthesized using a descriptive narrative review approach.

For KQ1 on comparative accuracy, we had planned to synthesize sensitivity and specificity of each test on its own using a bivariate mixed-effects binomial regression model as described by Harbord et al.¹⁷ All such analyses were to be computed by the STATA 13 statistical software package using the “midas” command.¹⁸ To compare two tests, we used the EPC methods described in equation 39 of Trikalinos et al. (2013).¹⁹

For KQ2 on comparative effectiveness, if meta-analysis was appropriate, we computed effect sizes and measures of variance using standard methods and performed Knapp-Hartung random-effects meta-analysis. If heterogeneity was encountered, we used meta-regression to explore possible causes.

For KQ3 on patient factors and tumor characteristics, we planned to calculate for each study the association between a patient factor (e.g., age) and comparative accuracy or effectiveness (e.g., the difference in sensitivity between two tests). If appropriate, we meta-analyzed the correlations using standard techniques and Knapp-Hartung standard errors.

We refrained from specifying a numerical threshold for clinical significance of results in either sensitivity or specificity in this report, for two reasons. First, there is no agreement in the field on what threshold is appropriate. Second, clinical importance is a clinical judgment that is sensitive to numerous clinicians and patient factors.

Grading the Body of Evidence for Each Outcome

We determined the strength-of-evidence grade for the following outcomes:

- Comparative sensitivity (KQ1)
- Comparative specificity (KQ1)
- Comparative timeliness of treatment (KQ2)
- Comparative choice of treatment (KQ2)
- Comparative tumor response (KQ2)
- Comparative survival (KQ2)
- Comparative quality of life (KQ2)
- Comparative harms due to overtreatment (KQ2)
- Comparative harms due to undertreatment (KQ2)
- Association between age and comparative accuracy (KQ3)
- Association between body habitus and comparative accuracy (KQ3)
- Association between tumor characteristics and comparative accuracy (KQ3)
- Association between age and comparative effectiveness (KQ3)
- Association between body habitus and comparative effectiveness (KQ3)
- Association between tumor characteristics and comparative effectiveness (KQ3)

We determined each grade separately for each modality comparison (e.g., CT vs. PET/CT). For accuracy, we examined both sensitivity and specificity. We did not separately grade other accuracy-related outcomes (i.e., predictive values or likelihood ratios) because these grades would be redundant with the grades for accuracy since our estimates would be based on the same studies and the same quantitative syntheses.

We used the EPC system for grading evidence on diagnostic tests as described in the EPC guidance chapter by Singh et al. (2012).²⁰ This system uses up to eight domains as inputs (study limitations, directness, consistency, precision, publication bias, dose-response association, all plausible confounders would reduce the effect, and strength of association). The output is a grade for the strength of evidence: high, moderate, low, or insufficient (see Table 1). This grade is provided separately for each outcome of each comparison of each KQ. The grade refers to our confidence in the direction of effect when comparing two imaging technologies, not to the magnitude of the difference between technologies.

Table 1. Strength of evidence grades and definitions

| Grade | Definition |
|----------|---|
| High | We are very confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies. We believe that the findings are stable, that is, another study would not change the conclusions. |
| Moderate | We are moderately confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains. |

| Grade | Definition |
|--------------|--|
| Low | We have limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect. |
| Insufficient | We have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion. |

Source: Singh et al. (2012).²⁰

A grade of Insufficient was given when the evidence did not permit a conclusion for the outcome of interest and the two modalities being compared. For example, if the outcome is test accuracy and the comparison is CT to PET/CT, the evidence may permit a conclusion that either 1) CT is more accurate, 2) PET/CT is more accurate, or 3) the tests are similarly accurate. If none of these three conclusions could be drawn (as judged by three independent analysts), then evidence was graded Insufficient for that comparison. In order to conclude that two modalities differ for an outcome, we used a p value less than 0.05 two-tailed (i.e., the standard value for alpha). In order to conclude that two modalities are approximately equivalent for an outcome, we used the independent judgment of three analysts (with disagreements resolved by discussion).

If the evidence was sufficient to permit a conclusion, then the grade was deemed high, moderate, or low. The grade was provided by three independent raters, with discrepancies resolved by consensus. Specifically, each of three analysts considered all strength of evidence domains listed earlier, and decided on a rating for each evidence base, without knowledge of the ratings of the other two analysts. If any of the three ratings differed, a single rating was reached based on consensus discussion.

Below, we discuss the eight domains and how they were considered as inputs to the grade:

Study Limitations (see the section Assessment of Methodological Risk of Bias of Individual Studies above). If the evidence permits a conclusion, and all else being equal, a set of studies with low limitations yield a higher strength of evidence grade than a set of studies with moderate or high limitations.

Directness. Our initial inclusion criteria required direct comparisons between modalities, but we subsequently included studies of single modalities. Indirect comparisons meant downgrades to the strength of the evidence. This is because different sets of studies may have enrolled slightly different types of patients or used slightly different reference standards for determining patients' true SCLC stages.

Consistency. Consistency among study results was judged based on whether the studies' findings suggest the same direction of effect (e.g., that one test is better than another).

Precision. Precision of the combined study results was judged by an assessment of the meta-analytic CI, or if meta-analysis was not performed, by an assessment of the precision of the individual studies (based on CIs or numbers of patients).

Reporting bias. This was addressed by noting the presence of abstracts or ClinicalTrials.gov entries describing studies that did not subsequently appear as full published articles. If many such studies exist, this tends to decrease the strength of the evidence. We also considered the funding sources of the studies in determining the risk of reporting bias.

Dose-response association. This factor is used for studies of treatments and is not relevant for this topic.

All plausible confounders would reduce the effect. This domain acknowledged that a set of studies may be biased *against* finding a difference between two modalities, and yet the studies still found an important difference. Thus, if the studies had controlled for the confounders, the effect would have been even larger. This domain will generally increase the strength-of-evidence grade.

Strength of association. While the Cochrane Collaboration and other reviewers have objective criteria for defining this domain in the context of treatments, no such criteria exist for diagnostic test accuracy. Therefore, this domain was judged by EPC team members based on whether the strength of the effect (e.g., the extent of difference in accuracy between two tests) is so large that the potential study biases could not explain it. If true, this domain will generally increase the strength-of-evidence grade.

Peer Review and Public Commentary

Peer reviewers were invited to provide written comments on the draft report based on their clinical, content, or methodologic expertise. Peer-review comments on the preliminary draft will be considered by the EPC in preparation of the final draft of the report. Dispositions of the peer-review comments are documented and will be published 3 months after publication of the evidence report.

Results

Results of Literature Searches

We depict the literature selection process in Figure 2. Searches identified 2,066 citations, of which we excluded 1,851. The most common reasons for exclusion were studies of other conditions (e.g., non-small-cell lung cancer), studies of treatments, and other studies not addressing staging. We retrieved the other 215 articles, of which we excluded 208. The most common reasons were studies with fewer than 10 patients with SCLC and studies of other conditions. See Appendix B for a list of the publications excluded at the full article level. We included the remaining seven publications. Our search of Clinicaltrials.gov identified no additional relevant studies.

All seven studies were included for KQ1, and none were included for KQs 2 or 3. Two studies were conducted in South Korea, and one each in Japan, Taiwan, Spain, Germany, and Denmark. The only study not published in English was one from Spain.²¹ The studies enrolled a total of 408 patients with SCLC, with average ages ranging from 56 to 68, and the percentage of patients who were female ranged from 15% to 62%. Start dates ranged from 1999 to 2008, and the median duration of patient enrollment was about 5 years. Four studies were retrospective, and three were prospective. Only one study²² reported its funding source (the German Cancer Foundation, which likely would not have a vested interest in some imaging technologies over others), and two other studies^{23,24} reported that there were no financial conflicts of interest to disclose.

All evidence tables appear in Appendix C, including general study information (Table C-1), patient characteristics (Table C-2), general test details (Table C-3), readers and reference standards (Table C-4), concordance data (Table C-5), accuracy data (Table C-6), and our analyses of accuracy data (Table C-7). Of the seven studies, three²³⁻²⁵ reported the comparative accuracy of two or more tests, and four^{21,22,26,27} reported single-test accuracy. One²⁵ of the comparative accuracy studies also reported concordance data.

Regarding the imaging tests performed:

- Three studies^{24,25,27} reported data on “standard staging” or “conventional staging.” This involved multiple tests, typically history, clinical exam, chest x-ray, bone marrow biopsy, and possibly MRI or CT of the brain.
- Three studies^{22,23,25} reported data on whole-body bone scintigraphy with Tc99m-MDP, using a dose of 500-900 MBq 2 to 3 hours before the scan, and a dual-head camera
- Two studies^{22,25} reported data on MDCT, but authors provided few details.
- Four studies^{21,23-25} reported data on PET/CT using FDG (dose range 400–550 MBq). One was specifically of the brain,²¹ and the others three were presumably whole body. All were dedicated PET/CT scanners: two were the GE Discovery system, one was the Phillips Gemini scanner, and one was the Siemens Biograph Sensation.
- One study²⁶ reported data on EBUS, using a convex probe integrated with convex transducer; a 22-gauge needle was used for needle aspiration.
- No studies reported data on MRI, PET/MRI, or EUS.

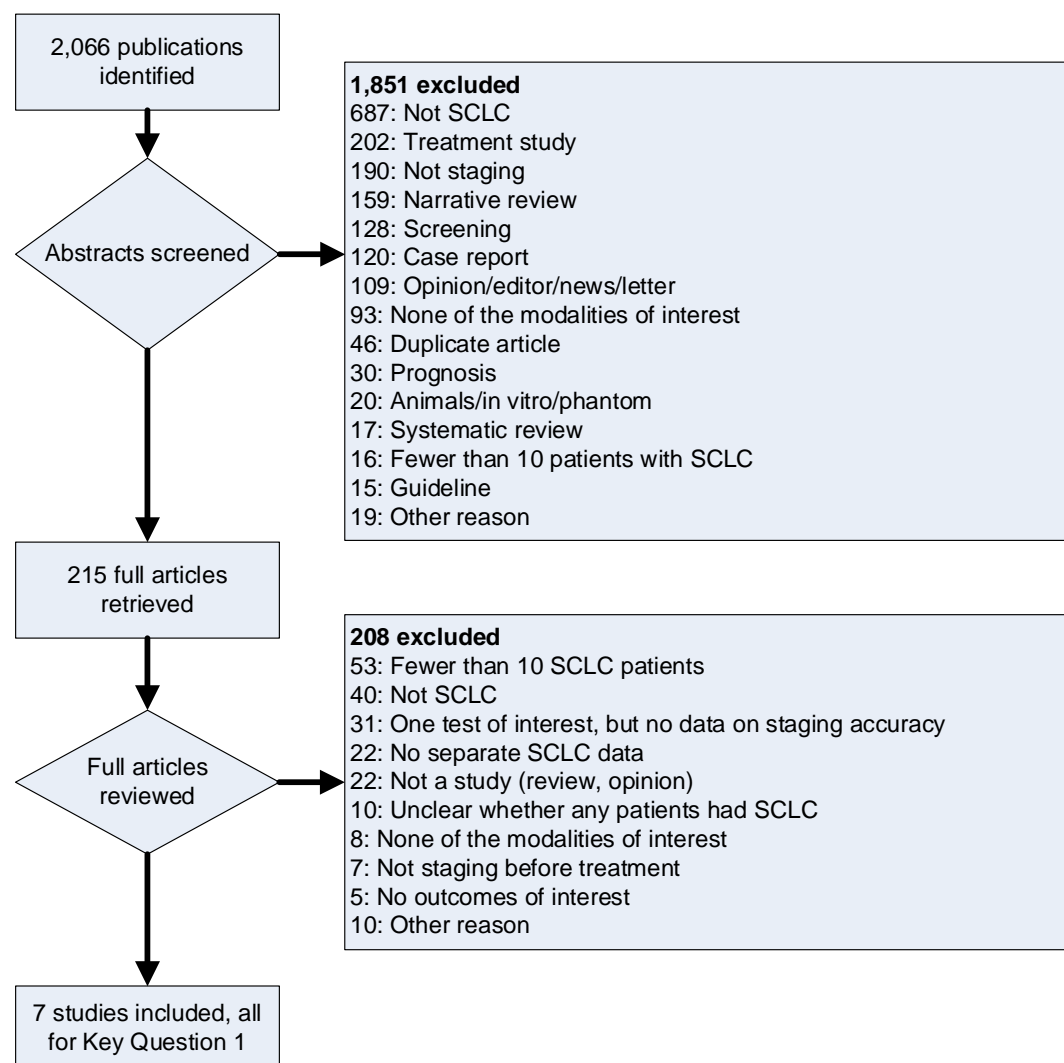
For three studies of comparative accuracy of imaging tests, the elapsed time between the tests was a median of 5 days in one study,²³ a maximum of 7 days in another study,²⁵ and not reported by the third study.²⁴ Test readers were generally experienced, but authors mentioned little about

resolving difference among different readers. Reference standards were based on multiple considerations such as histology if available, clinical followup, progression on subsequent scans, and other imaging results such as MRI.

Our risk-of-bias assessments appear in Appendix D in Figure D-1 and Figure D-2. Of the three studies reporting comparative accuracy, we rated three as moderate risk of bias and one as high risk-of-bias. The moderate ratings were due to unknown spectrum bias, lack of reporting of whether test readers had the same clinical information available when interpreting different tests, and the use of test results in determining the reference standard. The one high-risk-of-bias rating was due to the above problems as well as the probability of spectrum bias, not reporting the elapsed time between imaging tests, and non-complementary knowledge available to different test readers.

The four single-test accuracy studies were all rated as moderate risk of bias. Reasons for the moderate rating varied across studies, but common problems were not accounting for inter-reader differences and not blinding the reference standard assessment to test results or other clinical information.

Figure 2. Literature flow diagram



Key Question 1: Concordance and Comparative Accuracy

In this section, we address the following KQ:

KQ1: What are the test concordance and comparative accuracy of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) for the pretreatment staging of SCLC?

We first briefly summarize test concordance data, and then the rest of the section concerns comparative accuracy.

Test Concordance

One study²⁵ reported test-test concordance data for three imaging tests (MDCT, FDG PET/CT, and bone scintigraphy). The data appear in Table C-5 of Appendix C. For various staging determinations (T stage, N stage, pleural effusion, metastases to ipsilateral lung, metastases to contralateral lung, metastases to the liver, metastases to the adrenal glands, and

metastases to extra-thoracic lymph nodes), authors reported high agreement between MDCT and FDG PET/CT, ranging from 86 to 97 percent. For the assessment of osseous involvement, however, agreement was lower (83% between MDCT and FDG PET/CT; 46% between MDCT and bone scintigraphy; 57% between FDG PET/CT and bone scintigraphy). The same study also reported the accuracy of these modalities for assessing osseous metastases, and these data are discussed (along with all other accuracy data) in the next section.

Comparative Accuracy

An overview of the included accuracy data appears in Table 2. Studies reported numerous staging determinations (e.g., whether the patient has limited or extensive disease), but the evidence for any given determination and modality comparison was limited. The largest evidence base involved the comparison of FDG PET/CT to bone scintigraphy in the determination of osseous (bone or bone marrow) metastases; this evidence base comprised two studies making direct comparisons (combined n=123) and a single study reporting only bone scintigraphy accuracy data (n=76).

Below, we discuss the results separately for each of eight staging determinations (LD/ED, metastases to osseous structures (bone or bone marrow), lymph node involvement, metastases to adrenal glands, metastases to the liver, metastases to the spleen, any distant metastases, and metastases to the brain).

Table 2. Overview of included accuracy data

| Staging Determination | Studies Making Direct Comparisons in Accuracy Between Imaging Modalities | Studies Reporting Accuracy Data on a Single Imaging Modality (i.e., for indirect comparisons) |
|--|---|---|
| Limited or extensive disease | Standard staging vs. FDG PET/CT; 1 study; n=28 | Standard staging only; 1 study, n=25 |
| Presence of metastases to osseous structures (bone or bone marrow) | MDCT vs. Bone scintigraphy; 1 study; n=28 MDCT vs. FDG PET/CT; 1 study, n=29 Bone scintigraphy vs. FDG PET/CT; 2 studies, n=123 | Bone scintigraphy only; 1 study, n=76 |
| Presence of lymph node involvement | None | MDCT only; 1 study, n=118 EBUS only; 1 study, n=36 |
| Presence of metastases to adrenal glands | None | MDCT only; 1 study, n=120 |
| Presence of metastases to the liver | None | MDCT only; 1 study, n=120 |
| Presence of metastases to the spleen | None | MDCT only; 1 study, N=120 |
| Presence of any distant metastases | Standard staging vs. standard staging plus FDG PET/CT; 1 study; N=73 | None |
| Presence of metastases to the brain | None | FDG PET/CT only; 1 study, N=21 |

FDG PET/CT=[¹⁸F]-fluorodeoxyglucose positron emission tomography/computed tomography; MDCT=multidetector computed tomography

Limited or Extensive Disease

Two studies^{25,27} with moderate risk of bias reported data on the ability of imaging tests to determine whether patients with SCLC had LD or ED. Both reported the use of “standard

staging,” which is a combination of multiple testing procedures such as chest x-ray, bone marrow biopsy and possibly MRI or CT of the brain. In addition, Fischer et al. (2006)²⁵ reported data on the performance of FDG PET/CT in staging disease as LD or ED. The data suggest that both standard staging and FDG PET/CT performed well (e.g., 86% or 95% for the sensitivity at detecting ED, or specificity of 90% or more at ruling out ED). However, both studies were small (n=28 and n=25, respectively), and the overall data were too imprecise to permit any conclusions about relative accuracy.

Metastases to Osseous Structures (Bone or Bone Marrow)

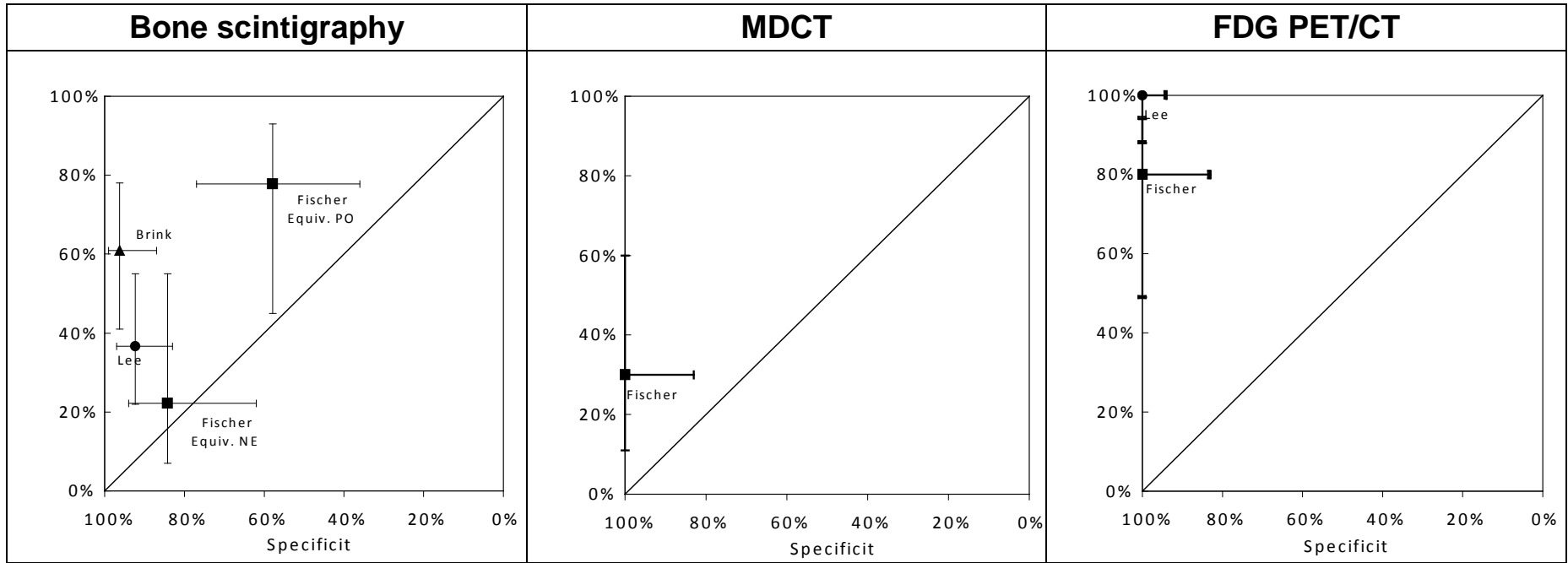
Three studies with moderate risk of bias reported data on the ability of imaging tests to determine whether patients had metastases to osseous structures (bone or bone marrow). One study compared FDG PET/CT to bone scintigraphy;²³ another study compared CT to FDG PET/CT to bone scintigraphy;²⁵ the third study provided data only on bone scintigraphy.²² The sensitivity and specificity data are shown in Figure 3. For bone scintigraphy, the data from Fischer et al. (2006)²⁵ are shown twice: once if equivocal bone scans are treated as positive tests and another if equivocal bone scans are treated as negative tests.

Our statistical tests of these data indicated FDG PET/CT was more sensitive than bone scintigraphy in the Lee study; the Fischer study replicated this finding if equivocal bone scans were treated as negative tests. If they were considered positive tests, then for the Fischer study FDG PET/CT was more specific than bone scintigraphy. These are direct comparisons. Considered together with the bone scintigraphy results in the Brink study, we concluded that FDG PET/CT is more sensitive than bone scintigraphy for detecting osseous metastases.

Comparing FDG PET/CT to MDCT, only the Fischer study made a direct comparison, and FDG PET/CT was more sensitive, but not more specific. Thus, we concluded that FDG PET/CT is more sensitive than MDCT for detecting osseous metastases. Too little data exist to permit a conclusion on specificity, so we cannot rule out FDG PET/CT being less specific than MDCT.

Finally, with regard to comparing MDCT with bone scintigraphy, treating equivocal bone scans as positive resulted in a statistical advantage in sensitivity for bone scintigraphy but a statistical advantage in specificity for MDCT. By contrast, if equivocal bone scans were treated as negatives in Fischer, we found no statistical differences in sensitivity or specificity. Other data on bone scintigraphy from Brink and Lee do not suggest marked differences from MDCT. However, the data were too imprecise to permit any conclusions about the comparison of MDCT and bone scintigraphy with respect to osseous metastases.

Figure 3. ROC plots of accuracy data for osseous metastases



Note: The three plots show the data on osseous metastases in ROC space. The left plot is for bone scintigraphy, the middle plot is for MDCT, and the right plot is for FDG PET/CT. In each plot, the horizontal axis is specificity (with higher specificity as one moves to the left in the plot), and the vertical axis is sensitivity (with higher sensitivity as one moves up in the plot). The 45 degree line is chance. Thus, an optimal modality would have data in the upper left corner of the plot. Each point is a study (with different studies represented by different shapes). The error bars represent 95% confidence intervals. The study by Brink provided data only on bone scintigraphy, which explains why it does not appear in the other two plots. The study by Lee compared bone scintigraphy and FDG- PET/CT, and the study by Fischer compared all three modalities. The Fischer data are further complicated by the fact that some bone scintigraphy results were considered equivocal by the authors (i.e., neither clearly positive nor clearly negative). Treating equivocal results as positive (as shown with the point labeled “Fischer Equiv. POS”) results in higher sensitivity than treating equivocal bone scan results as negative (as shown with the point labeled “Fischer Equiv. NEG”), but lower specificity. For PET/CT, both studies had point estimates for specificity of 100%, but the confidence intervals around those estimates were too wide to permit conclusions about relative specificity.

Lymph Node Involvement

Two moderate-risk-of-bias studies reported data on the ability of imaging tests to determine whether patients had lymph node involvement. One study used EBUS for this purpose;²⁶ the other used MDCT.²² The EBUS study reported better results (96% sensitivity, 100% specificity) than the CT study (70% sensitivity, 94% specificity). However, patients in the CT study may have had cancers that were more difficult to assess for lymph node involvement, which would bias the comparison against CT. The indirect nature of the comparison precludes conclusions.

Metastases to Adrenal Glands, Liver, or Spleen

A single study²² with moderate risk of bias reported the single-test performance of MDCT for detecting metastases to the adrenal glands, liver, or spleen (separate results for each of these three types). Because such data were not reported for other imaging modalities by this or other studies, we drew no conclusions about how different modalities compare.

Any Distant Metastases

A single study²⁴ with high risk of bias reported the comparative performance of standard staging versus standard staging plus FDG PET/CT for detecting any distant metastases. The study reported a large difference in sensitivity (92% for standard staging plus FDG PET/CT vs. only 46% for standard staging), and this difference was statistically significant. The specificities were similar (96% for standard staging plus FDG PET/CT vs. 100% for standard staging), but the precision was too high to permit a conclusion of equivalence on specificity. We deemed the evidence sufficient to permit the conclusion that standard staging plus FDG PET/CT is more sensitive than standard staging alone for detecting any distant metastases. Given that it was only a single study that had a high risk of bias, we rated the strength of evidence as low.

Metastases to the Brain

A single study with moderate risk of bias reported the single-test performance of FDG PET/CT of the brain for detecting brain metastases. Because such data were not reported for other modalities by this or other studies, we drew no conclusions about how different modalities compare in the assessment of brain metastases.

Key Question 2: Comparative Effectiveness

In this section, we address the following KQ:

KQ2: When used for the pretreatment staging of SCLC, what is the comparative effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) on later outcomes?

No studies were included for this question.

Key Question 3: Factors Associated with Comparative Outcomes

In this section, we address the following KQ:

KQ3. To what extent are the following factors associated with the comparative accuracy or effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) when used for the pretreatment staging of SCLC?

No studies were included for this question. No studies had been excluded due to the fact that we required such comparisons to be planned in advance.

Discussion

Key Findings and Strength of Evidence

Based on the evidence we reviewed, we concluded the following:

- FDG PET/CT is more sensitive than bone scintigraphy at detecting osseous metastases (Strength of Evidence: low)
- FDG PET/CT is more sensitive than CT at detecting osseous metastases (Strength of Evidence: low)
- Standard staging plus FDG PET/CT is more sensitive than standard staging alone at detecting any distant metastases (Strength of Evidence: low)

Our strength-of-evidence judgments for these conclusions, along with the other evidence that was insufficient to permit conclusions, are listed in Table 3 below. Note that all three of our conclusions involve the superior sensitivity of FDG PET/CT. We discuss this commonality below in the sections “Finds in Relationship to What Is Already Known” and “Implications for Clinical and Policy Decisionmaking.”

Table 3. Strength of evidence grades

| Staging Determination | Test Comparison | Number of Studies and Number of Patients | Study limitations | Direct-ness | Consistency | Precision | Reporting Bias | Magnitude of Effect | SOE Grade* | Conclusion |
|------------------------------|----------------------------------|--|-------------------|-------------|-------------|-----------|----------------|---------------------|--------------|---------------------------|
| LD vs. ED | Standard staging vs. FDG PET/CT | Direct comparison: 1 study ²⁵ of 28 patients. Indirect comparison: 1 study ²⁷ of standard staging of 25 patients | Moderate | Mixed | Unknown | Imprecise | None suspected | Not large | Insufficient | NA |
| Osseous metastases | FDG PET/CT vs. bone scintigraphy | Direct comparison: 2 studies ^{23,25} of 123 patients. Indirect comparison: 1 study ²² of bone scintigraphy of 76 patients | Moderate | Mixed | Consistent | Imprecise | None suspected | Large | Low | FDG PET/CT more sensitive |
| Osseous metastases | CT vs. FDG PET/CT | Direct comparison: 1 study ²⁵ of 29 patients | Moderate | Direct | Unknown | Imprecise | None suspected | Large | Low | FDG PET/CT more sensitive |
| Osseous metastases | CT vs. bone scintigraphy | Direct comparison: 1 study ²⁵ of 28 patients. Indirect comparison: 1 study ²² of bone scintigraphy of 76 patients | Moderate | Mixed | Unknown | Imprecise | None suspected | Not large | Insufficient | NA |
| Lymph node involvement | CT vs. EBUS | Indirect comparison: 1 study ²² of CT of 118 patients, and 1 study ²⁶ of EBUS of 36 patients | Moderate | Indirect | Unknown | Imprecise | None suspected | Not large | Insufficient | NA |
| Metastases to adrenal glands | CT vs. anything else | Indirect comparison: 1 study ²² of CT of 120 patients | Moderate | Indirect | Unknown | Unknown | None suspected | Unknown | Insufficient | NA |
| Metastases to liver | CT vs. anything else | Indirect comparison: 1 study ²² of CT of 120 patients | Moderate | Indirect | Unknown | Unknown | None suspected | Unknown | Insufficient | NA |
| Metastases to spleen | CT vs. anything else | Indirect comparison: 1 study ²² of CT of 120 patients | Moderate | Indirect | Unknown | Unknown | None suspected | Unknown | Insufficient | NA |

Table 3. Strength of evidence grades (continued)

| Staging Determination | Test Comparison | Number of Studies and Number of Patients | Study limitations | Direct-ness | Consistency | Precision | Reporting Bias | Magnitude of Effect | SOE Grade* | Conclusion |
|------------------------|---|---|-------------------|-------------|-------------|-----------|----------------|---------------------|--------------|---------------------------|
| Any distant metastasis | Standard staging vs. Standard staging plus FDG PET/CT | Direct comparison: 1 study ²⁴ of 73 patients | High | Direct | Unknown | Imprecise | None suspected | Large | Low | FDG PET/CT more sensitive |
| Metastases to brain | FDG PET/CT of the brain vs. anything else | Indirect comparison: 1 study ²¹ of FDG PET/CT of 21 patients | Moderate | Indirect | Unknown | Imprecise | None suspected | Unknown | Insufficient | NA |

EBUS=endobronchial ultrasound; ED=extensive stage disease; FDG PET/CT =[18F]-fluorodeoxyglucose positron emission tomography/computed tomography; LD=limited stage disease; NA=Not applicable since evidence was insufficient to permit a conclusion for this staging determination for this test-test comparison; SOE=strength of evidence

* The SOE grade indicates our confidence in the conclusion about the direction of the effect, not about the magnitude of the difference.

Findings in Relationship to What Is Already Known

Our searches found eight previous systematic reviews that evaluated imaging modalities for pretreatment staging of SCLC. We found no previous systematic reviews of evidence on CT, MRI, or bone scintigraphy for SCLC staging. Thus, our conclusions for these modalities (and documentation of lack of evidence in many instances) represent new knowledge that can guide future investigations.

All prior systematic reviews on imaging modalities for SCLC staging focused on PET. Some reviews included studies using combined PET/CT scanners, but most had data only from standalone PET scanning, which is no longer the current standard of practice. Our review focuses on PET/CT, which will more accurately reflect choices available today.

The most relevant previous reviews were undertaken in the development of American College of Chest Physicians (ACCP) guidelines for management of SCLC. Two editions of these guidelines included reviews meeting our inclusion criteria. In the review by Samson et al. (2007),²⁸ for the second edition of the guidelines, only one KQ addressed the role of imaging tests in SCLC staging. This question asked whether the addition of PET scanning improved staging of SCLC. The authors identified six studies of PET for various staging indications, deemed the evidence to be poor quality, and drew no conclusions from it. This and the previous reviews could not use a consistent reference standard for comparison to PET; they had to use the reference standards reported by the investigators of each study.

Since the publication of that review, combined PET/CT systems have replaced standalone PET scanners at many hospitals, so our review excluded standalone PET, which resulted in exclusion of all the studies Samson et al. cited. Like Samson et al, we also found the evidence on PET/CT to be weak, but nevertheless concluded that PET/CT has superior sensitivity compared to either CT alone or bone scintigraphy for detecting osseous metastases; we also concluded that PET/CT has superior sensitivity to standard staging for detecting any distant metastases overall at the patient level. Samson et al. saw a similar trend in the standalone PET studies they reviewed. We are more willing to state conclusions from the weak evidence because the AHRQ evidence grading system expresses the degree of weakness and makes clear that additional evidence could overturn our conclusions.

The third edition of the ACCP guidelines included the KQ “what is the ability of PET imaging to determine the stage of cancer?” The review concluded that PET was superior to standard staging modalities for detecting metastases, with the exception of brain metastases, where PET was inferior to CT and MR. The resulting guideline recommended PET for patients with SCLC clinically diagnosed as limited-stage.

While the review for ACCP by Jett et al. (2013)¹⁰ was published relatively recently, it included studies dating back as far as 2001; nearly all the studies used standalone PET. Direct integration of anatomic data from CT in combined PET/CT may mitigate some inaccuracy of PET in diagnosing brain metastases. Indeed, when looking only at PET/CT compared to standard staging, we could not conclude that PET/CT was inferior.

The Jett review also analyzed how often PET resulted in a change of management plans for patients with SCLC. All but one study used standalone PET, but a substantial number of patients remained whose management was changed as the result of PET findings. Like we did, Jett et al. recognized that the absence of a reliable reference diagnosis in most of these studies precludes determination that PET is the superior imaging modality. Jett et al. also commented on the lack

of uniformity in study methods and data analysis; these issues persist in the evidence base for PET/CT and hindered our ability to draw any conclusions.

Neither of the ACCP guideline reviews offered a granular categorization and analysis of data such as we have provided. Therefore, although we were unable to draw many conclusions about the comparative effectiveness of PET/CT, we have provided a thorough and detailed description of the current available evidence that future investigators can utilize to address these important evidence gaps.

The two other systematic reviews of SCLC staging—Lu et al. (2014)²⁹ and Ruben et al (2012)³⁰—were similar in that the majority of studies in both were small and included standalone PET. However, the focus of the two reviews differed. Lu et al. sought to determine a summary sensitivity and specificity of PET for distinguishing between patients with LD and ED, concluding that the sensitivity of PET was 98.1 percent (95% CI 94.7%–99.6%) and the specificity was 97.5 percent (95% CI 93.0%–99.5%). We did not find sufficient data to permit similar summary estimates for PET/CT. Ruben et al. studied the effect of PET on patient management, concluding that changes to treatment plans occurred in nearly one-third of patients having PET scans. Jett et al. arrived at a similar conclusion. In the absence of a gold standard reference, it is not possible to ascertain whether those changes resulted in better or worse outcomes for the patient due to denial of potentially valuable treatment (false-positive PET) or avoidance of unnecessary treatment and its side effects (true-positive PET).

Three previous reviews—Ravanal (2012),³¹ Helwig et al. (2009),³² and Ung et al. (2007)³³—covered lung cancer staging in general, making little reference to SCLC in particular. As the clinical course of SCLC is considerably different from that of other forms of lung cancer, decisions on staging and treatment should be based on evidence specific to SCLC. The effects of differences between cancer types could outweigh the differences in effectiveness between one modality and another as seen in a study of mixed types of lung cancer. The systematic review published by the U.K. National Health Service on PET for various cancer indications³⁴ reported that as of the time of their searches (August 2005) no published studies existed on combined PET/CT for diagnosis, staging, or restaging of SCLC.

Implications for Clinical and Policy Decisionmaking

Based on our review of the current evidence, our results suggest two overall conclusions. First, compared to CT and bone scintigraphy (imaging modalities commonly used for staging) FDG PET/CT is more sensitive for detecting osseous metastases in patients with SCLC. Our findings suggest that clinicians evaluating patients for the presence of osseous metastases may consider forgoing bone scintigraphy and routinely use FDG PET/CT instead. Second, for patients with SCLC who have undergone standard staging, the addition of FDG PET/CT increases sensitivity for detecting any distant metastases overall at the patient level.

The evidence base did not allow us to draw conclusions about the comparative specificity of FDG PET/CT compared to these other modalities; thus, we acknowledge that any of the modalities could yield false positives. If a false positive led to inaccurate “upstaging” to extensive disease, a patient might only receive a palliative regimen instead of aggressive chemotherapy aimed at cure. Unfortunately, as we found only 1 study addressing EBUS and no studies evaluating EUS, MRI, or PET/MRI for inclusion we were unable to assess their comparative accuracy for FDG PET/CT.

SCLC is an aggressive cancer, and timely staging is important to determine treatment decisions based on whether patients have limited or extensive disease. Currently, as part of the

standard staging process, patients may undergo bone scintigraphy, CT of the abdomen and pelvis, brain MRI, and FDG-PET/CT. Reducing the total number of tests may improve the timeliness of staging and permit faster initiation of treatment.

Higher sensitivity also has other potential important implications for patient care. First, better detection of metastases can improve patient selection for optimal therapy. The higher sensitivity of FDG PET/CT would result in accurate determination of extent of disease, providing clinicians the confidence to offer a comprehensive stage-based treatment plan. Second, earlier detection of extensive disease would spare patients from more aggressive chemotherapy and radiation protocols used for patients with LD. Earlier initiation of palliative measures may result in improved quality of life, an important consideration given the current poor prognosis of this disease. Third, improved sensitivity and timeliness of staging may potentially improve the ability of ongoing research trials to prognosticate and detect therapeutic efficacy.

Finally, our results suggest potential resource implications. Although FDG PET/CT may be more expensive than CT and bone scintigraphy, some patients may undergo all three tests when evaluating patients for osseous and other distant metastases outside the brain. Eliminating routine use of bone scintigraphy from SCLC staging protocols in favor FDG-PET/CT could potentially result in some cost savings.

Applicability

We judged the applicability of the evidence based on comparison of the patients, interventions, and settings found in the research studies to those typically used in clinical practice. The typical mean age of patients in our included studies was in the seventh decade of life. Based on the National Cancer Institute's (NCI) Surveillance Epidemiology and End Results (SEER) database, in 2011, 82 percent of those living with lung cancer were 60 years of age or older.³⁵ The majority of studies involved a predominance of male patients (3-5:1 male/female ratio) and one involved a slight female predominance (1:1.5 male/female ratio). Based on the NCI SEER database, in 2011, men develop lung cancer more often than women (1.3:1 male/female ratio).³⁵ Overall, the applicability of patients in the included studies appears reasonable.

With regard to imaging tests, the studies of FDG-PET/CT employed patient preparation and acquisition parameters typical of those used in clinical practice,³⁶ despite minor variations in image-acquisition parameters used in these included studies. Similarly, bone scintigraphy performed in these studies was obtained with techniques typical of those used in clinical practice.³⁷ EBUS-FNA is also frequently used as a diagnostic test for staging SCLC, and the one study addressing this technique used it similarly to that used in clinical practice.

Studies were conducted in university-based academic or teaching hospitals outside the United States, which may limit the applicability of the results to community hospitals and hospitals in the United States. Community hospitals may differ from the settings in the included studies with respect to the experience of the technicians administering the imaging test or the interpretation skills of those reading the imaging results. U.S. hospitals may differ from those outside the United States in terms of their specified protocols for the diagnostic workup and management of patients with SCLC.

Limitations of the Comparative-Effectiveness Review Process

In this section, we discuss three challenges we faced in developing this review: (1) whether to include the concept of diagnosis, (2) whether to include non-English studies, and (3) how to assess the risk of bias of comparative accuracy studies.

The first challenge arose from the fact that imaging tests often serve the dual purpose of diagnosing and staging lung cancer. By definition, when imaging is used for diagnosis, it remains unknown whether the patient has SCLC. Lung symptoms may be due to a variety of causes, including SCLC, but also NSCLC or metastases from another site, in addition to noncancerous causes. As all these etiologies are possibilities during the process of diagnosis, including the concept of diagnosis would have required including studies of imaging tests for diagnosing any of these possibilities. Given the defined time and budget, such a scope was untenable, and so we chose to focus on the staging of SCLC.

A second challenge involved whether to include non-English studies. We initially excluded them due to the cost of translation and the possibility that the design and results of such studies may not be applicable to the United States. However, when confronted with the small size of the resulting literature (5 studies), we reconsidered the 28 studies that we had excluded for this reason. Of those, two studies met our inclusion criteria.

A third challenge involved assessing the risk of bias of comparative accuracy studies. The basic target for this assessment is whether a study comparing the staging accuracy of test A to that of test B (measuring both against a common reference standard) was biased in favor of one of the two tests. Ideally, we could have used an existing off-the-shelf assessment instrument. However, current risk-of-bias instruments for diagnostic studies (e.g., QUADAS-2) do not sufficiently address this topic because they were designed for single-test accuracy studies (e.g., did this study provide unbiased estimates of test accuracy). We thought carefully about potential areas of bias and devised our own instrument for this purpose. The instrument has not been tested by others, and its appropriateness should be verified.

Limitations of the Evidence Base

The seven included studies were limited in numerous ways. The most prominent ways were (1) lack of test-test comparisons, (2) small sample sizes, (3) questionable reference standards, and (4) lack of comparative-effectiveness designs.

First, only three studies directly compared the accuracy of different imaging tests. More test-test comparisons might have permitted us to draw more conclusions from the evidence. Indirect comparisons are fraught with difficulty because important inter-study differences in patient populations could confound any comparisons.

Second, the seven studies were generally small (the median sample size was 36, with a range from 21 to 120). This means large imprecision in statistical results may result, which in turn means insufficient evidence to make claims about test-test comparisons.

Third, studies' reference standards were often suboptimal. Studies generally utilized a combination of clinical history, cell pathology (only obtained for a portion of patients), other imaging tests, and clinical followup. These reflect the clinical reality of the process of staging SCLC, since it would be unethical to perform a biopsy on all possible sites of metastases in a patient. Thus, the test accuracy estimates reported by all studies are tentative. The direction of possible bias resulting from the lack of good reference standards in these studies is not

predictable, so it simply increases the uncertainty of the results. For instance, incorporation of the study test into the reference standard (incorporation bias) would tend to increase positive predictive value, but other shortcomings such as lack of histologic verification of positives could decrease positive predictive value.

Fourth, and perhaps most important, none of the studies were designed to allow inferences about *comparative effectiveness*. Generally, all patients in a given study received the same battery of imaging tests; thus, subsequent outcomes (e.g., management strategies, survival time and quality of life) could not be attributed to any single test or test strategy. The clinical value of an imaging test is best measured by using it for only some of the patients and comparing their health outcomes to a carefully matched group of other patients who did not receive that imaging test. Granted, such a design is exceedingly rare in the medical literature. Without such designs, however, the true patient-oriented value of any give staging modality cannot be determined.

Research Gaps

For characterizing gaps, we used the Hopkins EPC framework proposed by Robinson et al. (2011).³⁸ That system suggests that reviewers identify a set of important gaps and determine the most important reason for each gap. Each gap should be assigned one of the following reasons for the inability to draw conclusions:

- A. Insufficient or imprecise information: no studies, limited number of studies, sample sizes too small, estimate of effect is imprecise
- B. Information at risk of bias: inappropriate study design; major methodologic limitations in studies
- C. Inconsistency or unknown consistency: consistency unknown (only 1 study); inconsistent results across studies
- D. Not the right information: results not applicable to population of interest; inadequate duration of interventions/comparisons; inadequate duration of followup; optimal/most important outcomes not addressed; results not applicable to setting of interest

We found three primary gaps in the literature on imaging tests for the pretreatment staging of SCLC. The first concerns the dearth of evidence on several tests of interest, particularly MRI, EBUS, EUS, and PET/MRI. This gap exists due to reason A, Insufficient information. Note that we did not restrict our search to studies of particular staging purposes of these tests even though some are typically used for specific targets (e.g., brain MRI). EBUS and EUS may be used as much for diagnosis as for staging, so that may partially explain the lack of direct evidence on staging SCLC. PET/MRI is a relatively new technology, and we predicted that little would be identified, but future systematic reviews may uncover evidence as it becomes more widespread.

The second gap concerns the absence of study designs to inform the optimal imaging plan for the pretreatment staging of SCLC. Thus, the reason for this gap is reason B, Information at risk of bias. At least three underlying sources exist for the gap: (1) the general lack of direct comparisons of different imaging modalities; (2) the use of mixed reference standards (based on partial histology, other imaging tests, and clinical follow-up) since not all patients' true stage can be determined before treatment; and (3) the complete lack of studies of comparative effectiveness with patient-oriented outcomes such as management strategy or survival after receiving different test strategies. We recognize that some of these problems are due to the clinical reality that SCLC is fast-growing, necessitating timely initiation of treatment.

The third gap concerns KQ3, which addressed the extent to which comparative accuracy or effectiveness are associated with patient factors (comorbidities, body habitus, tumor characteristics). We identified no studies for this question, so this gap exists due to reason A, Insufficient information. Addressing this problem would be easier than addressing the problems listed in the previous paragraph. As most patient records already contain information on patient characteristics, future research could stratify their accuracy results accordingly. Armed with this more detailed data, clinicians and policymakers could possibly recommend tailoring specific strategies for different patient subgroups.

References

1. American Cancer Society (ACS). Cancer facts & figures 2014. Atlanta (GA): American Cancer Society (ACS); 2014. 70 p.
<http://www.cancer.org/acs/groups/content/40research/documents/webcontent/acspc-042151.pdf>.
2. SEER Cancer Statistics Review, 1975-2011. Bethesda (MD): National Cancer Institute; 31 p.
3. American Cancer Society. Cancer prevention & early detection facts & figures 2013. Atlanta (GA): American Cancer Society; 2013. 65 p.
4. National Cancer Institute (NCI). Small cell lung cancer treatment (PDQ) health professional version. Bethesda (MD): National Cancer Institute (NCI), National Institutes of Health (NIH); 2014 Feb 21.
<http://www.cancer.gov/cancertopics/pdq/treatment/small-cell-lung/healthprofessional/page1>. Accessed 2014 Jun 11.
5. Shepherd FA, Crowley J, Van Houtte P, et al. The International Association for the Study of Lung Cancer Lung Cancer Staging Project: proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. *J Thorac Oncol*. 2007 Dec;2(12):1067-77. PMID: 18090577
6. Kalemkerian GP, Akerley W, Bogner P, et al. Small cell lung cancer. *JNCCN J Natl Compr Canc Netw*. 2013 Jan 1;11(1):78-98. PMID: 23307984
7. Micke P, Faldum A, Metz T, et al. Staging small cell lung cancer: Veterans Administration Lung Study Group versus International Association for the Study of Lung Cancer--what limits limited disease? *Lung Cancer*. 2002 Sep;37(3):271-6. PMID: 12234695
8. Trends in lung cancer morbidity and mortality. Chicago (IL): American Lung Association; 2012 Apr. 34 p.
9. Glisson BS, Byers LA. Pathobiology and staging of small cell carcinoma of the lung. In: Lilenbaum RC, Nicholson A, Jett JR, editors. *UptoDate*. Waltham (MA): UptoDate; 2013 May 07.
http://www.uptodate.com/contents/pathobiology-and-staging-of-small-cell-carcinoma-of-the-lung?source=search_result&search=sclc&selectedTitle=2%7E150#H7. Accessed 2014 Jun 12.
10. Jett JR, Schild SE, Kesler KA, et al. Treatment of small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013 May;143(5 Suppl):e400S-19S. PMID: 23649448
11. Hillner BE, Tosteson AN, Song Y, Et al. Growth in the use of PET for six cancer types after coverage by medicare: additive or replacement? *J Am Coll Radiol*. 2012 Jan;9(1):33-41. PMID: 22221634
12. 2014 PET market summary report. Greenbelt (MD): IMV Info; 2014.
<http://www.imvinfo.com/index.aspx?sec=pet&sub=dis&itemid=200083>. Accessed 2015 May 15.
13. New IMV survey shows slowing growth in PET medical imaging scans. Des Plaines (IL): PRWeb; 2012 Aug 12.
<http://www.prweb.com/releases/2012/8/prweb9820176.htm>. Accessed 2015 May 18.
14. Lang K, Huang H, Lee DW, et al. National trends in advanced outpatient diagnostic imaging utilization: an analysis of the medical expenditure panel survey, 2000-2009. *BMC Med Imag*. 2013;13:40. PMID: 24279724
15. Ravenel JG, Rosenzweig KE, Kirsch J, et al. ACR appropriateness criteria non-invasive clinical staging of bronchogenic carcinoma. *J Am Coll Radiol*. 2014 Sep;11(9):849-56. PMID: 25086958
16. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003 Nov 10;3(1):25. PMID: 14606960

17. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007 Apr;8(2):239-51. PMID: 16698768
18. STATA statistics/data analysis. MP parallel edition. Ver. 10.0. College Station (TX): StataCorp; 1984-2007. Single user Stata for Windows. <http://www.stata.com>.
19. Trikalinos TA, Hoaglin DC, Small KM, et al. Evaluating practices and developing tools for comparative effectiveness reviews of diagnostic test accuracy: methods for the joint meta-analysis of multiple tests. Methods research report. (Prepared by the Tufts Evidence-based Practice Center, under Contract No. 290-2007-10055-I.) AHRQ publication no. 12(13)-EHC151-EF. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2013 Jan. 49 p. http://effectivehealthcare.ahrq.gov/ehc/products/291/1380/Methods%20Report_Evaluating-Practices-Developing-Tools_Final_01-14-2013.pdf.
20. Singh S, Chang S, Matchar DB, et al. Grading a body of evidence on diagnostic tests (AHRQ publication no. 12-EHC079-EF). Chapter 7 of Methods guide for medical test reviews (AHRQ publication no. 12-EHC017). Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2012 Jun 1. http://effectivehealthcare.ahrq.gov/ehc/products/246/558/Methods-Guide-for-Medical-Test-Reviews_Full-Guide_20120530.pdf.
21. Palomar Munoz A, Garcia Vicente AM, Bellon Guardia ME, et al. Is a selective brain (18)F-FDG PET/CT study profitable in patients with small cell lung cancer? *Rev Esp Med Nucl Imagen Mol*. 2012 May;31(3):124-9. PMID: 21722995
22. Brink I, Schumacher T, Mix M, et al. Impact of [(18)F]FDG-PET on the primary staging of small-cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2004 Dec;31(12):1614-20. PMID: 15258700
23. Lee JW, Lee SM, Lee HS, et al. Comparison of diagnostic ability between (99m)Tc-MDP bone scan and 18F-FDG PET/CT for bone metastasis in patients with small cell lung cancer. *Ann Nucl Med*. 2012 Oct;26(8):627-33. PMID: 22729551
24. Sohn BS, Lee DH, Kim EK, et al. The role of integrated 18F-FDG PET-CT as a staging tool for limited-stage small cell lung cancer: a retrospective study. *Onkologie*. 2012 Aug;35(7):432-8. PMID: 22846975
25. Fischer BM, Mortensen J, Langer SW, et al. A prospective study of PET/CT in initial staging of small-cell lung cancer: comparison with CT, bone scintigraphy and bone marrow analysis. *Ann Oncol*. 2007 Feb;18(2):338-45. PMID: 17060487
26. Wada H, Nakajima T, Yasufuku K, et al. Lymph node staging by endobronchial ultrasound-guided transbronchial needle aspiration in patients with small cell lung cancer. *Ann Thorac Surg*. 2010 Jul;90(1):229-34. PMID: 20609781
27. Shen YY, Shiau YC, Wang JJ, et al. Whole-body 18F-2-deoxyglucose positron emission tomography in primary staging small cell lung cancer. *Anticancer Res*. 2002;22(2):1257-64. PMID: 12168935
28. Samson DJ, Seidenfeld J, Simon GR, et al. Evidence for management of small cell lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*. 2007 Sep;132(3 Suppl):314S-23S. PMID: 17873177
29. Lu YY, Chen JH, Liang JA, et al. 18F-FDG PET or PET/CT for detecting extensive disease in small-cell lung cancer: a systematic review and meta-analysis. *Nucl Med Commun*. 2014 Jul;35(7):697-703.
30. Ruben JD, Ball DL. The efficacy of PET staging for small-cell lung cancer: a systematic review and cost analysis in the Australian setting. *J Thorac Oncol*. 2012 Jun;7(6):1015-20. PMID: 22534816
31. Ravenel JG. Evidence-based imaging in lung cancer: a systematic review. *J Thorac Imaging*. 2012 Sep;27(5):315-24. PMID: 22914125
32. Hellwig D, Baum RP, Kirsch C. FDG-PET, PET/CT and conventional nuclear medicine procedures in the evaluation of lung cancer: a systematic review. *Nuklearmedizin*. 2009;48(2):59-69. PMID: 19333516

33. Ung YC, Maziak DE, Vanderveen JA, et al. 18Fluorodeoxyglucose positron emission tomography in the diagnosis and staging of lung cancer: a systematic review. *J Natl Cancer Inst.* 2007 Dec 5;99(23):1753-67. PMID: 18042932
34. Facey F, Bradbury I, Laking G, et al. Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers. *Health Technol Assess.* 2007 Oct;11(44):iii-85. PMID: 17999839
35. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review (CSR) 1975-2011. Bethesda (MD): National Cancer Institute (NCI); 2014 Dec 17. http://seer.cancer.gov/csr/1975_2011/. Accessed 2015 Feb 2.
36. American College of Radiology (ACR), Society for Pediatric Radiology (SPR). ACR-SPR practice parameter for performing FDG-PET in oncology. Reston (VA): American College of Radiology (ACR); 2014. 14 p. <http://www.acr.org/~media/71B746780F934F6D8A1BA5CCA5167EDB.pdf>.
37. American College of Radiology (ACR), Society for Pediatric Radiology (SPR). ACR-SPR practice parameter for the performance of skeletal scintigraphy (bone scan). Reston (VA): American College of Radiology (ACR); 2014. 9 p. <http://www.acr.org/~media/839771405B9A43F7AF2D2A9982D81083.pdf>.
38. Robinson KA, Saldanha IJ, Mckoy NA. Frameworks for determining research gaps during systematic reviews. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2011 Jun. 79 p. (Methods future research needs report; no.2). http://www.effectivehealthcare.ahrq.gov/ehc/products/201/735/FRN2_Frameworks_20110726.pdf.

Abbreviations and Acronyms

| | |
|-----------|---|
| 95% CI | 95% confidence interval |
| CT | Computed tomography |
| EBUS | Endobronchial ultrasound |
| EUS | Endoscopic ultrasound |
| FDG | Fluorodeoxyglucose |
| MBq | Megabecquerel |
| MHz | Megahertz |
| MRI | Magnetic resonance imaging |
| NA | Not applicable |
| NR | Not reported |
| PET/CT | Combined positron emission tomography and computed tomography |
| QUADAS | Quality assessment tool for diagnostic accuracy studies |
| SCLC | small cell lung cancer |
| SD | Standard deviation |
| TBNA | transbronchial needle aspiration |
| Tc99m-MDP | Technetium-99m methylene diphosphonate |